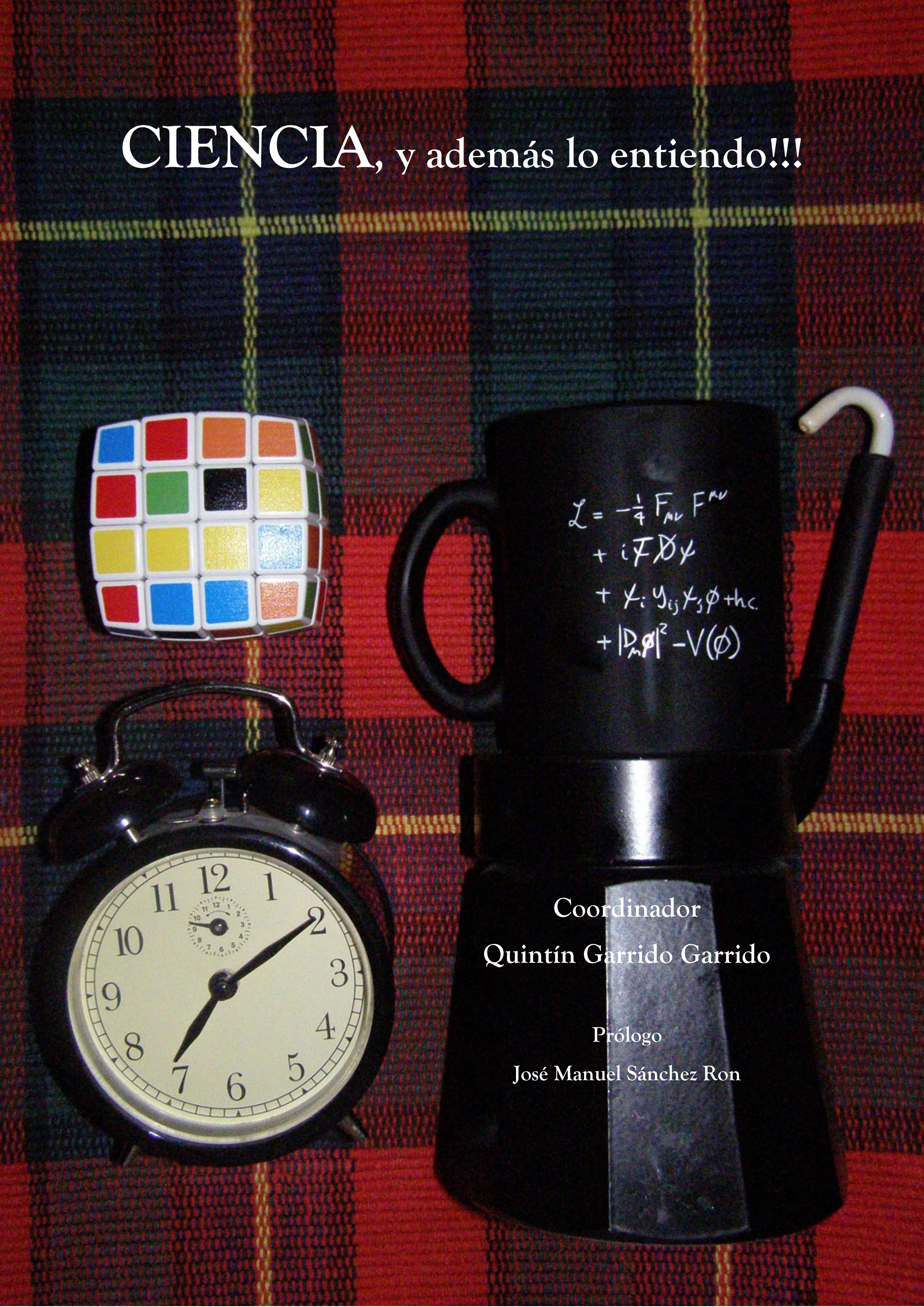


CIENCIA, y además lo entiendo!!!

Coordinador

Quintín Garrido Garrido

Prólogo

José Manuel Sánchez Ron

CIENCIA, y además lo entiendo!!!

Este libro tiene una licencia Creative Commons
Cualquier reproducción total o parcial de esta obra deberá hacer un reconocimiento expreso a la autoría de la misma y/o de los capítulos mencionados.
No se permite el uso comercial de la obra original ni de las posibles obras derivadas.
La distribución y uso de las obras derivadas se debe hacer bajo una licencia igual a la que se regula la obra original.
Las imágenes, figuras, ilustraciones que aparecen están amparadas bajo esta licencia salvo en las que expresamente se hace mención a su autoría/crédito al pie de las mismas.



Presentación y agradecimientos.

Este libro surge, como se indica en el comienzo de ese email que ha llegado a multitud de científicos de todo el mundo, de *"mi penúltima idea feliz"*.

Que mejor manera para celebrar el 2º aniversario de un blog que recoge, fundamentalmente, los libros de divulgación científica recomendados por los propios científicos, que con la elaboración de un libro, de divulgación claro está.

Y ya puestos un pequeño tributo a uno de los grandes del género, Isaac Asimov y en concreto a su famoso "100 Preguntas básicas sobre la Ciencia".

Para esta ingente labor he tenido el honor de poder contar con la colaboración de (lo mejor de cada casa) grandes científicos de las más variadas disciplinas, de multitud de centros docentes y/o centros de investigación, tanto públicos como privados. Todos ellos con una gran trayectoria y con un futuro realmente muy prometedor. Este libro, y mi mayor agradecimiento, es de todos vosotros, yo únicamente me he limitado a aglutinaros entorno a este proyecto y a "bombardearos" con emails, plazos, etc. Gracias por acogerme entre vosotros y por involucraros de una manera seria, responsable y hacer posible llevar a cabo todo esto.

Agradecer a todos los que, por diversos motivos, no habéis podido participar. Vuestro ánimo y buenos deseos para con este proyecto también han ayudado a llevarlo a delante.

Gracias a Pedro y a Pablo por su ayuda con el "Inglés", sin vuestro apoyo y traducciones no hubiera sido lo mismo.

Gracias a Alberto y a José Manuel, vuestro apoyo, ayuda e interés han sido fundamentales para seguir hacia delante.

Mención especial para el autor del primer capítulo, D. Ricardo, mi profesor de Física del Instituto y, lo supe después de que se animaran a participar, de varios de los autores. Gran profesor, gran persona y gran amigo. Gracias por todo.

Y unas palabras de agradecimiento, las más importantes, para David, Pablo y Teresa. Gracias por vuestro apoyo y por permitirme dedicarle parte del tiempo familiar a este hobby, "**CIENCIA"**.

Con la esperanza de que los lectores de este libro disfrutéis tanto al leerlo como todos nosotros al elaborarlo. Y también que la próxima "Idea Feliz" sea tan agradable de realizar y llevar a buen puerto como esta.

¡No dejéis nunca de tener la imaginación y curiosidad de un niño!

Madrid Noviembre 2016
Un Abrazo
Quintín

Prólogo

Difundir los conocimientos científicos, la ciencia, es una de las tareas más nobles que conozco, especialmente si, como sucede en el presente caso, lo hace, de manera completamente desinteresada, un numeroso grupo de científicos, coordinados – más correcto sería decir, liderados – por Quintín Garrido Garrido. Digo que es una de las tareas más nobles que conozco, porque la ciencia es, de lejos, el mejor instrumento que han creado los humanos para librarse de mitos, de esas muy abundantes ideas que no son sino fruto de la imaginación, que, por supuesto, puede obedecer a razones muy variadas, entre las que sin duda se hallan algunas perfectamente comprensibles dada la naturaleza humana, que busca seguridad y permanencia. De la imaginación, y no pocas veces también de intereses particulares. Sin la ciencia no podemos entendernos a nosotros mismos, ni a todo lo que nos rodea, el medio, terrestre y cósmico, que llamamos bien Naturaleza o Universo. Es cierto que, al menos por el momento, la ciencia no proporciona respuestas a todas las preguntas que podemos imaginar – ¿quién sabe si lo logrará alguna vez?; yo lo dudo –, pero cada día da alguna respuesta nueva, y aunque no suministrara más a partir de ahora, cosa que no sucederá, ¡menudo equipaje nos ofrece ya!

Acabo de referirme a las preguntas que contesta la ciencia, y precisamente es, en general, recurriendo a preguntas – en campos muy variados: matemáticas, física, astrofísica, química, biología, neurociencias, geología, oceanografía, ciencia de los materiales… – cómo está estructurado el presente libro, este *Ciencia, y además lo entiendo!!!* El esfuerzo que sus autores han realizado porque se entienda lo que escriben es digno del mayor reconocimiento. No existe mayor peligro, mayor enemigo de la ciencia que acorralarla en la oscuridad de presentaciones técnicas, especializadas. Es evidente que ese tipo de presentaciones son las propias de la dinámica de la investigación científica, pero es absolutamente necesario salir en ocasiones de ese mundo tan cerrado en sí mismo. Aunque la sociedad pueda no reclamarlo – desgraciadamente, esto sucede con frecuencia –, lo necesita. Y cuando esa sociedad, la ciudadanía, recibe explicaciones claras y amenas de lo que es la ciencia y sus contenidos, lo agradece.

No ignoro que algunas de las respuestas que brinda la ciencia seguramente no nos harán felices. Como Darwin y sus seguidores nos enseñaron, no somos, ay, el fruto privilegiado de un Creador todopoderoso, sino polvo de estrellas que se condensó, dando origen a muy diferentes formas de vida, mediante procesos no dirigidos de prueba y error, que, eso sí, obedecieron a las leyes que va desvelando la ciencia. El azar de caminos en los que reinó lo fortuito, y la necesidad de cumplir lo que imponen las leyes naturales. Pero si la ciencia no da siempre felicidad, sí que da dignidad. Entre los atributos que más admiro de los humanos, se encuentra el ser capaces de actuar noblemente no teniendo la esperanza de *la eternidad*, siendo conscientes, muy conscientes, de nuestra contingencia.

Como Quintín señala en su presentación, este libro celebra el segundo aniversario de un blog que se ocupa de libros de divulgación científica, y aunque quien

escribe estas líneas no frecuente demasiado tales lugares del hiperespacio digital – por demasiado apego a una galaxia, la Gutenberg, en vías de desaparecer, cual si fuese una supernova a punto de explotar –, no ignoro sus muchas virtudes, entre las que destacan el no conocer fronteras, lo que significa estar a disposición de cualquiera, y la generosidad de quienes dedican parte de su tiempo a componerlos. Como este libro, es un ejemplo de generosidad y de creencia en la importancia de lo común, que es importante celebrar. Al igual que la aparición de esta benemérita obra.

José Manuel Sánchez Ron
De la Real Academia Española

Índice

Capítulo 1
¿Qué es el método científico?
(Por Ricardo David Fernández Cruz)

Tal vez la primera cuestión a preguntarnos es ¿qué es la Ciencia? Una respuesta simple sería, la actividad que hacen los científicos, sin embargo tal contestación, nos deja en el mismo estado de desconocimiento que teníamos anteriormente. La Ciencia Moderna es una de las actividades humanas, que se ocupa de descubrir el comportamiento de muchos de los fenómenos que ocurren en el mundo, ¿para qué? Podemos dar varias razones:

-Para satisfacer la curiosidad que tenemos los seres humanos de conocer los fenómenos naturales, pero, a diferencia de otros saberes, para la adquisición de este conocimiento, se utiliza un procedimiento particular, conocido como “el método científico”. Con él, los científicos realizan en general, el descubrimiento y la descripción de los fenómenos, utilizando como principal medio de trabajo la experimentación. Además se emiten hipótesis, y se desarrollan leyes y teorías, que tienen generalmente un carácter temporal. Los conocimientos designados como teorías científicas, no tienen naturaleza eterna e inmutable, y están sometidos de continuo a revisión, lo que permite su confirmación, remodelación o sustitución. En ocasiones, para que una teoría de una respuesta suficiente a ciertas observaciones, es necesario ampliarla con algunos ajustes. Cuando ya no pueda responder a las nuevas observaciones y nuevos datos experimentales, ha de ser sustituida por otra teoría más amplia, que incluye como caso particular a la anterior. *La Ciencia es como un edificio, en continua y perpetua construcción, donde las nuevas teorías se van asentando encima de las anteriores, que le van sirviendo de soporte.*

-Para emplear el conocimiento científico adquirido, en beneficio de la humanidad, como una fuente de riqueza, mejorando nuestra salud, calidad de vida y disminuir el esfuerzo en el trabajo. Esta finalidad se realiza mediante la aplicación de los conocimientos científicos a fines prácticos, a través de la Tecnología. En este apartado cabe señalar, que en la actualidad y a lo largo de la Historia, se encuentran ejemplos que muestran, como los conocimientos científicos no siempre han sido aplicados para el bien de la humanidad. Este uso desafortunado, generalmente no es responsabilidad de los científicos, sino de ciertos grupos de presión, que en ocasiones obligan a determinados científicos a realizar aplicaciones tecnológicas, con finalidades poco deseables. Existen numerosos ejemplos en la Historia de la Ciencia, de científicos que se han negado a realizar aplicaciones prácticas de su saber, que pudieran contribuir a perjudicar o incluso a la destrucción de seres humanos.

-Para seguir proporcionando un mayor y mejor conocimiento del universo, -consideramos con esta palabra el estudio de cualquier objeto, sea grande o pequeño, próximo o lejano-. Se conoce esta actividad, como investigación básica, operación que

nunca debería detenerse, aunque necesite inversiones económicas importantes, pues sin duda es la herramienta que nos seguirá abriendo el progreso en el futuro.

Finalmente queremos señalar, que aunque el trabajo científico debe ser una actividad libre, pues sin libertad no hay creación, sin embargo, la sociedad tiene el derecho de conocer y la obligación de opinar sobre el trabajo científico, por varias razones: por ser el soporte económico de las investigaciones, y por motivos mucho más profundas de índole ético, ya que no se puede dejar en manos de un colectivo, por importante que éste sea, la realización de actividades que puedan dañar o comprometer, la dignidad de los seres humanos.

Mediante el método científico, los científicos procuran construir con precisión, las representaciones de los fenómenos que observamos en la naturaleza. En él se tienen en cuenta las influencias que pueden tener los prejuicios, las predilecciones personales y la cultura, en nuestra percepción e interpretación del mundo. El método científico utiliza procedimientos establecidos para minimizar estas inclinaciones del experimentador, cuando ensaya hipótesis o teorías. El método científico se organiza en varios pasos, sin embargo, hay una gran libertad tanto en su aplicación, como en su número. Aquí vamos a considerar cuatro etapas:

1) Observación y descripción del fenómeno o fenómenos que se quieren llegar a interpretar.

2) Formulación de una hipótesis para explicarlo.

3) Empleo de la hipótesis para predecir la existencia de nuevos fenómenos o los resultados de nuevas observaciones.

4) Realización de pruebas experimentales para verificar las predicciones, por varios experimentadores y mediante procedimientos distintos.

Si experimentalmente se confirman las hipótesis, pueden ser consideradas como una teoría o ley de la naturaleza. Si el experimento no confirma las hipótesis, deberán ser modificadas o abandonadas. Está rigurosamente establecido, que las pruebas experimentales tienen la primacía sobre cualquier otra, para la confirmación de las hipótesis, o en la decisión sobre una cierta hipótesis. El método científico requiere que una hipótesis sea plenamente confirmada por otros experimentadores, y cuando sucede que las predicciones son incompatibles con las pruebas experimentales, ésta hipótesis debe ser modificada. Los experimentadores pueden verificar directamente la teoría, o confirmarla a través de consecuencias derivadas de ella, usando las matemáticas y la lógica, y en todas las ciencias empíricas el experimento es el juez supremo, siendo de absoluta necesidad para la verificación de las hipótesis. Aquellas teorías que no puedan ser probadas mediante observaciones medibles, no pueden ser calificadas como científicas.

Una clave del método científico, es su poder predictivo, sin embargo estas predicciones deben ser comprobadas por la observación y el experimento. Se suele decir con frecuencia "que las teorías nunca son probadas, solo desaprobadas". Siempre hay la posibilidad de que una nueva observación, entre en conflicto con una teoría planamente

establecida, cuando esto sucede, pongamos como ejemplo la mecánica clásica desarrollada por Newton en la segunda mitad del siglo XVII, y se deducen predicciones teóricas que están en desacuerdo con nuevos resultados experimentales, la teoría debe ser descartada como una descripción completa de la realidad, pero puede continuar aplicándose dentro de ciertos límites. Así, las leyes de la mecánica clásica tienen validez cuando las velocidades son muchos menores que la luz en el vacío y en ese caso, (al que se adaptan la mayor parte de nuestras experiencias cotidianas), dichas leyes se utilizan ampliamente, tanto en la ciencia como en la tecnología. Sin embargo, cuando las velocidades se aproximan a la de la luz, o dentro de intensos campos gravitatorios, los fenómenos se describen mejor mediante la Teoría de la Relatividad, desarrollada por Einstein a comienzos del siglo XX, de la que se encontraron pruebas experimentales de su validez, con posterioridad.

Las hipótesis y las teorías científicas, deben ser comprobadas experimentalmente, y en esta verificación, influyen tanto el experimentador, como los instrumentos de medida. De un modo general, se determina el grado de confianza en una medida mediante el cálculo de las incertidumbres experimentales, existiendo varias fuentes de error. Están el error intrínseco del instrumento de medida, que tiene igual probabilidad de producirse con valores más altos y bajos, del valor verdadero (que es desconocido) y el error sistemático, debido a factores, que desvían los datos en un sentido determinado. En ciencia, se disponen de procedimientos establecidos de estimación de incertidumbres, lo que resulta imprescindible para calcular la precisión de una medida particular, y cuando se determinan resultados cuantitativos, para acotarla. Una medida sin una cota de incertidumbre resulta inaceptable.

El método científico intenta minimizar la influencia sesgada de los propios científicos en los resultados de sus experimentos, porque cuando comprueban una hipótesis, pueden tener cierta preferencia por un resultado u otro, y es importante que estas no condicionen el resultado o su interpretación. Un error especialmente grave consiste en la no elaboración de una hipótesis para la explicación de un fenómeno, cuando se realizan pruebas experimentales, pensando que se trata de un paso innecesario, porque los resultados deben ser los que predicen el sentido común o la lógica, y otro error muy común es ignorar o buscar alguna explicación no comprobada, para eliminar aquellos datos que no se adaptan a la hipótesis.

En todo experimento, la hipótesis inicial puede ser correcta o incorrecta. Pero, a veces, los científicos tienen una gran confianza en que una hipótesis es verdadera o falsa, o se sienten presionados para conseguir determinados resultados. En estos casos, puede haber una predisposición a encontrar justificaciones y aceptar datos que coincidan con las expectativas deseadas, con posibles efectos sistemáticos. Para evitarlo, todos los datos deben ser tratados de la misma forma.

Hay ejemplos de descubrimientos, desechados por los experimentadores, con información de nuevos fenómenos, pero que únicamente fueron considerados en último término, y a la inversa, existen casos de pretendidos nuevos descubrimientos, que más tarde se comprobó que provenían de errores sistemáticos, no contemplados por sus descubridores. En los campos en que hay una experimentación activa y comunicación entre los científicos, los sesgos individuales o de un grupo, se cancelan, porque los experimentos son realizados por personas distintas y con medios diferentes, que como es de esperar tendrán distintas tendencias.

Mediante el método científico se elaboran hipótesis, modelos, teorías y leyes:

Una hipótesis es una explicación limitada, que contempla una causa y su efecto en situaciones muy concretas, siendo emitida con nuestro conocimiento del fenómeno, antes de que el trabajo experimental haya sido ejecutado. Tomando un ejemplo de la vida diaria, supongamos que una lámpara no luce. Una primera hipótesis sería, "la bombilla está fundida" y podemos comprobar su validez, cambiándola por otra nueva. Si continúa sin lucir, la hipótesis no era cierta, y plantearemos una segunda hipótesis; "el interruptor está averiado". Para comprobarla hay que sustituirlo por otro, o verificarlo con un polímetro. Así, poco a poco, emitiendo hipótesis y rechazando aquellas que no sean certeras, conseguiremos emitir una que finalmente permita solucionar el problema.

Un modelo es una representación más sencilla y asequible a nuestro conocimiento, de un fenómeno real. A partir de un modelo se elaboran las teorías científicas o leyes, cuya validez se corresponde con la del modelo elegido. En el modelo de Bohr para el átomo de hidrógeno los electrones describen órbitas circulares alrededor del núcleo, y ésta no es una correcta descripción del átomo, sin embargo es sencillo matemáticamente, y explica razonablemente muchas características del átomo.

Una teoría científica o ley**,** representa una hipótesis o un grupo de hipótesis relacionadas, las cuales se han visto confirmadas repetidas veces mediante pruebas experimentales. Las teorías físicas son normalmente formuladas con un reducido número de conceptos y ecuaciones, las cuales son consideradas como leyes de la naturaleza, insinuándose su aplicación universal. Una vez aceptadas, se convierten en herramientas para mejorar nuestro conocimiento del universo, permitiendo explorar áreas desconocidas situadas en la frontera de nuestro saber. Las teorías no son fáciles de descartar, de forma que los nuevos descubrimientos tratan de encajarse dentro de la teoría existente, y solo sí después de repetidas pruebas experimentales, el nuevo fenómeno no puede ser acomodado, los científicos cuestionan seriamente la teoría e intentan su modificación. Los cambios en la Ciencia y en las teorías ocurren por supuesto, dando paso a teorías nuevas, que modifican nuestra visión del mundo. Nuevamente "la fuerza para el cambio" la proporcionan el método científico y la experimentación.

Sin embargo el método científico no es aplicable a todos los sucesos de la vida cotidiana, en aquellos casos en los que no sea posible aislar los fenómenos, para verificar las hipótesis, o cuando las medidas no puedan ser repetidas nuevamente, el método científico no es de aplicación. En la bolsa de valores, oímos constantemente que éstos suben y baja sin unas causas claras que lo justifiquen, aquí el método científico (por desgracia para los científicos pues ganarían mucho dinero) no es de aplicación. Existen innumerables causas que influyen en la cotización de las acciones, que no pueden ser aisladas una por una, además no se puede experimentar con los valores, haciéndolos subir y bajar a voluntad, para conocer la respuesta de los inversores y así deducir leyes que controlen el comportamiento de las acciones. Algo análogo ocurre en los juicios, los abogados no pueden repetir el delito delante del juez o del jurado, para que juzguen a la vista de las pruebas.

Un papel fundamental de la ciencia actual es la comunicación y la publicidad. Todos los científicos que realizan un trabajo de interés, tanto teórico como

experimental, lo difunden, mediante su publicación en alguna de las revistas especializadas que existen, en papel o a través de la red. Solo de esta forma, se hace posible que otros verifiquen la validez de sus resultados o de las hipótesis y teorías propuestas. Se considera como verdad científica, siempre provisional, la que una vez publicada, resulta aceptada por un número suficiente de expertos en el tema. Excepto en casos poco frecuentes, como proyectos de investigación militar, la información científica está abierta a todos los que puedan estar interesados y además debe ser comunicada a la sociedad, que en definitiva es la que la sustenta económicamente y la destinataria final de la misma.

Ricardo David Fernández Cruz
Doctor en Ciencias Física
Catedrático de Física de Instituto, jubilado.

Capítulo 2
¿Por qué dos o más científicos sin conocer el trabajo de otros, dan a menudo simultáneamente con la misma teoría?

(Por José Luis Rubio)

El tema parece sorprendente, pero no lo es tanto. En realidad, muchos inventos, avances y descubrimientos se han producido simultáneamente por individuos o grupos de investigación que trabajaban independientemente. La historia está llena de casos famosos. Quizás uno de los más conocidos sea el Darwin y Wallace con el descubrimiento simultáneo e independiente de la teoría de la evolución. Mientras Charles Darwin se debatía en su agónica lucha interna por dar a conocer o no lo que él sabía que iba a ser una auténtica revolución, Alfred Russel Wallace, trabajando independientemente en la lejana Nueva Zelanda, había llegado a las mismas conclusiones. De hecho si Wallace, honesta y confiadamente, no hubiera enviado a Darwin sus escritos, quizás hoy conoceríamos la teoría de la evolución como la teoría de Wallace y no como la teoría de Darwin.

Lo que parece ser una casualidad se repite muchas veces a lo largo del desarrollo científico y en los avances tecnológicos. Existen abundantes recopilaciones y listados de descubrimientos completamente independientes, sean sincrónicos o no, a lo largo de la historia de la humanidad. Robert K. Merton (1973) es uno de los autores más conocidos en el estudio de los descubrimientos simultáneos a lo largo del tiempo. Por cierto que uno de los avances científicos que cita es el descubrimiento de la circulación pulmonar de Miguel Servet (1553) que se produjo con total independencia del mismo descubrimiento por Ibn al-Nafis en Egipto (1242).

En el mundo de las patentes y del desarrollo tecnológico, a lo largo de la historia, nos encontramos con sonoras y dramáticas luchas, no solo por la autoría, sino por la prioridad de los descubrimientos. Frecuentemente se han aireado disputas y acusaciones, algunas veces con demandas por plagio, espionaje y fraude, cuando en realidad se llegó al mismo resultado de manera independiente. Un caso muy llamativo es el de Alexander Bell y Elisha Gray que presentaron su solicitud de patente sobre el teléfono, ¡el mismo día!, en concreto el 14 de febrero de 1876. Es también famoso el caso de Tesla y Edison, con sus múltiples conflictos, en principio resueltos a favor del norte americano por una cierta falta de escrúpulos y mayor capacidad de presión y maniobra, pero paulatinamente con un mejor y justo reconocimiento de las contribuciones de Tesla.

Esta pauta de simultaneidad también puede observarse en los desarrollos de las culturas y civilizaciones a lo largo del tiempo. También, en los colectivos sociales sin contacto alguno, se han ido produciendo descubrimientos y avances simultáneos e independientes. Uno de los de mayor trascendencia quizás haya sido el de la agricultura, que surge casi sincrónicamente en zonas tan alejadas y evidentemente sin absoluto contacto entre ellas, como el Creciente Fértil-Mesopotamia, China, Mesoamérica y la

India. En este mismo sentido, sociedades aisladas de distintos rincones del mundo progresaron y descubrieron avances que también lograron civilizaciones lejanas. Estos avances se produjeron no solo en agricultura, sino en una variedad enorme de temas como pueden ser la escritura, el calendario, las matemáticas, la arquitectura, la organización social, el aprovechamiento del agua, el arte, la fabricación de herramientas y utensilios, la domesticación de animales,…

¿Por qué se produce esta situación? Los investigadores del tema consideran que más que casos únicos, se les debe considerar como una pauta común en ciencia y en el patrón del progreso humano (Kuhn, 1962). De alguna manera se va produciendo un proceso acumulativo de experiencias, observaciones y progresos, muchas veces basados en la práctica común de prueba y error. Se va creando un cierto clímax de época o de conocimiento acumulado que en un momento dado genera la irrupción del descubrimiento. Pero este proceso tiene múltiples facetas e implicaciones.

Existe una curiosa similitud entre el desarrollo y evolución del individuo y la evolución del colectivo humano y social. La creatividad siempre ha sido y me atrevería a decir, que todavía lo es, un proceso misterioso. En tiempos, se pensaba que la creatividad era algo ajeno al ser humano, algo externo que procedía de la inspiración de las "musas" o de poderes superiores externos. Algo que venía de no se sabía bien de dónde pero que era ajeno al individuo. Hoy sabemos que se trata de una actividad cognitiva que, como otras funciones mentales, la desarrolla nuestro sistema neuronal. Hemos pasado de considerarlo, como algo así como un favor de los dioses, a una función de nuestras neuronas (Lehrer, 2012). Pero estamos todavía en los inicios de entender el fascinante proceso del descubrimiento o de la creatividad y de momento, en su esencia, el hecho de encontrar la solución al problema planteado o el hacer algo de manera distinta, permanece como un gran misterio rodeado de falsos mitos.

En efecto y durante mucho tiempo se ha considerado al descubrimiento como algo inescrutable, como una cualidad biológica que solo alcanzaba a algunos pocos afortunados. Sin embargo, el ser humano sigue unas pautas comunes de comportamiento marcadas por nuestro sistema genético como especie. Dentro de estas pautas comunes algunos individuos logran avances y desarrollos que están fuera del patrón general y, además, esta situación puede ser compartida por otros individuos con los que no existe contacto de ningún tipo. En distintas individualidades, e individuos aislados, se llega a una situación de experiencia, conocimiento y atracción por lo no conocido o no experimentado, en la que el desencadenamiento del descubrimiento puede surgir de manera casi inevitable. Pero esta situación previa al descubrimiento deriva y es consecuencia del conocimiento previo acumulado. En este sentido la frase de Newton:

"Si he visto más lejos es porque estoy sentado sobre los hombros de gigantes"

ilustra perfectamente esta situación. Por cierto, se trata de una metáfora que se atribuye a Newton pero que, al parecer, no fue, ni mucho menos, el primero en utilizarla.

Así pues, puede llegarse a una situación de clímax cuyo paso siguiente es un eureka que previamente permanecía oculto. Y una vez alcanzado este clímax o momentun, si un investigador no consigue el descubrimiento, muy probablemente otro lo hará. Sin embargo de este nivel de clímax no se deriva necesariamente que el descubrimiento sea inevitable. La naturaleza no revela fácilmente sus secretos. La mayoría de los inventos tienen detrás un largo y lento proceso acumulativo que puede durar décadas y en el que progresivamente se han ido produciendo avances en distintos

aspectos necesarios y complementarios, que finalmente crean el momentun para que el individuo, o individuos aislados, puedan producir el descubrimiento. Se podría afirmar que, los avances se producen solo cuando llega ese "momento". Sin embargo a nivel individual y previamente al descubrimiento son necesarias ciertas condiciones. En primer lugar es necesaria una etapa previa de saturación de conocimientos sobre el tema. También es necesaria la capacidad de observación y de disposición mental para la evaluación adecuada de los hechos, datos y observaciones. De alguna manera, la mente humana solo ve las cosas para las que está preparada para ver. De ahí la oportuna frase de Louis Pasteur en la que advierte que "el azar favorece al espíritu preparado". Algunas veces, y si se han dado estos aspectos previos, puede surgir el fenómeno de serendipia, cuando nos hemos relajado y alejado mentalmente del tema.

Pero para que se produzca el descubrimiento o avance y que este pueda implementarse, también son necesarias otras condiciones externas al investigador, como pueden ser la capacidad social y económica de su entorno para hacerse eco del descubrimiento y desarrollarlo. También el contexto cultural e institucional que pueda apreciar el interés de su aplicación. Por supuesto, y como en todo, la suerte y la confluencia de situaciones favorables, pueden ser un factores importantes.

Cuando ocurren todas estas circunstancias, individuales y sociales, la invención puede producirse. Puede ocurrir lo que Henry James llamaba "adivinar lo invisible a partir de lo visible" o según Martin Heidegger, el "proceso de desocultación". Y este proceso o epifanía tiene lugar en nuestra red neuronal. Todavía nadie sabe como. Todavía no se conoce como puede producirse ese fogonazo de segundos de duración, en el que súbitamente aparece la comprensión del problema y la solución del mismo. Según los estudios neurológicos (Lehrer 2012), el fogonazo de ondas gamma procede de una circunvolución del lóbulo temporal superior derecho. Y una vez ocurrido, la solución del problema resulta obvia. ¿Pero cómo no se me había ocurrido? Ahora bien y como hemos indicado, hay que estar preparado para poder darnos cuenta de esa tenue y rápida ráfaga de corriente neuronal en la que fugazmente aparece la solución. Hace falta una importante preparación personal y unas adecuadas condiciones sociales. Si estas se dan, podremos darnos cuenta de ese ramalazo de tenues ondas neuronales en el que viaja la respuesta al problema o la invención largamente buscada. Es el eureka que solo durante décimas de segundo nos envían nuestras sinapsis. En definitiva, una nueva idea o una solución es una pauta neuronal que de pronto cambia y funciona con un nuevo patrón que previamente no se había producido hasta entonces.

Este proceso puede producirse en mentes alejadas trabajando independientemente en el mismo problema. Y también puede producirse en colectivos científicos o sociales trabajando conjuntamente. De hecho, hoy día se reconoce que las grandes contribuciones científicas surgen de la colaboración de grupos, y cada vez estas colaboraciones aumentan su carácter internacional y multidisciplinar. Existe una clara tendencia a la colaboración y al trabajo en equipo, en el desarrollo de la ciencia. Es significativa la pauta creciente a otorgar Premios Nobel a grupos de científicos más que a investigadores individuales. Existen razones de tipo práctico como es la necesidad de distintos especialistas que cubran distintos campos científicos y la necesidad de utilizar costosas y sofisticas infraestructuras de investigación.

Para terminar una reflexión-consejo. Dado que nos encontramos en un mundo cada vez más interconectado y con los antecedentes que hemos comentado, si tienes

algo interesante en tu cabeza o en tu laboratorio, corre a publicarlo o a patentarlo porque si no, es muy probable que alguien se te pueda adelantar.

Referencias:
Kuhn, T. S. 1962 La estructura de las revoluciones científicas. Fondo de Cultura Económica
Lehrer, J. 2012 Imaginar. Cómo funciona la creatividad. RBA.
Merton, R. K. 1973 "Resistance to the Systematic Study of Multiple Discoveries in Science", European Journal of Sociology, 4:237–82, 1963. Reprinted in Robert K. Merton, The Sociology of Science: Theoretical and Empirical Investigations, Chicago, University of Chicago Press,1973 (https://en.wikipedia.org/wiki/List_of_multiple_discoveries)

José Luis Rubio
Doctor Ingeniero Agrónomo
Investigador Científico CIDE – CSIC, Valencia

Capítulo 3
Infinito
(Por Enrique Zuazua)

"Hasta el infinito y más allá" suele decir el valiente, fortachón y simpático Buzz Lightyear, uno de los protagonistas de la película de animación Toy Story. Es una frase que uno solo puede pronunciar con convencimiento cuando es niño, personaje de cómic o, de adulto, en algún momento puntual de euforia pasajera, pues pronto aprendemos que, en la vida, hay pocas cosas infinitas o tal vez ninguna. De niños descubrimos que la vida es finita ya sea por la muerte de una mascota, de un miembro de la familia o de algún personaje famoso que no conocemos pero importante en nuestro pequeño universo. Recuerdo la preocupación que me produjo la muerte de Walt Disney en 1966. "No habrá más dibujos animados ni tebeos con sus personajes" pensé.

Y en realidad todo tiene necesariamente un horizonte finito pues nuestra propia vida tiene los días contados, pocos o muchos. A pesar de ello, tenemos una clara intuición del infinito con el que convivimos en diversos ámbitos.

Hay un infinito filosófico y místico. En la tradición cristiana, por ejemplo, después de esta vida hay otra que es eterna, infinita y en la que debemos creer, tener fe. Debemos así intentar ser buenos para que lleguemos a la nueva vida sin fin del más allá con pocas cuentas pendientes, pues cada una de ellas habrá de ser purgada y, a poco que el castigo o penitencia sea doloroso, al durar toda la eternidad, su peso se nos hará insoportable.

El infinito puede tener distintos colores; rojo como el infierno, azul como el cielo, o gris como el purgatorio, pero siempre es ilimitado. El Dios de nuestra tradición cristiana es también una representación de ese infinito. Diríamos que es una encarnación del infinito si no fuese porque es incorpóreo. Dios tiene infinitos poderes y bondad y vela por el orden dentro del caos en el que nos vemos envueltos, pues el Planeta Tierra no deja de ser un enorme y desorganizado hormiguero.

La necesidad de transcendencia, de dar a la vida humana una perspectiva de duración infinita, más allá de lo que conocemos en nuestra experiencia sobre este planeta, es un elemento recurrente en todas las civilizaciones. Ya nuestros clásicos filósofos griegos, Aristóteles, Platón, Pitágoras, concibieron la necesidad del infinito y analizaron sus posibles formas y las consecuencias que tendría y contradicciones que generaría la aceptación de su existencia. Pasaron más de dos mil años más hasta que tan profunda cuestión quedó bien cimentada. Pero hay también infinitos más cotidianos que se nos presentan en el día a día. Por ejemplo, la línea del horizonte en la que se encuentran el cielo y el mar nos parece que está en el infinito. Por mucho que nademos o volemos hacia ella nunca la alcanzaremos, siempre estará más allá. Y esa es precisamente otra de las características principales del infinito, es inalcanzable.

Infinito puede también ser el amor que una persona experimente hacia otra, hasta el punto de preferir sacrificar su propia vida a experimentar la pérdida de la de la otra.

Puede que el universo sea también infinito, pues vivimos en la Tierra, en el sistema solar, dentro de la galaxia de la Vía Láctea, que no es más que una entre otras muchas. Pero, ¿de dónde cuelga toda esa construcción? ¿dónde está clavada la chincheta que lo sujeta? Todo sería más fácil si el espacio fuese infinito pues entonces no tendríamos que preocuparnos de donde colocamos su principio ni su final.

El infinito es también uno de los conceptos centrales de las Matemáticas y en ellas hay numerosas paradojas que lo evocan, alguna incluso no exenta de moraleja. Es el caso de la famosa tortuga perezosa que experimentó su propio infinito. Tenía que visitar a su familia que vivía a una distancia de un kilómetro. Pero era tan perezosa que el primer día solo hizo la mitad del camino, medio kilómetro, el segundo día la mitad del recorrido que le faltaba por hacer, es decir un cuarto del camino total, el tercero un octavo pues era la mitad del cuarto que le quedaba por recorrer. Y así siguió un día tras otro hasta que se dio cuenta de que con ese plan de viaje nunca llegaría a su destino, pues siempre le faltaría por andar la mitad del día anterior y el doble del siguiente. Había conseguido que una distancia finita se convirtiera en infinita, imposible de alcanzar, como consecuencia de su infinita pereza. El infinito es obligado y ubicuo en Matemáticas, en efecto.

El ser humano inventó los números para contar y medir, lo cual era indispensable para el comercio, para construir y organizar ciudades,… Y al hacerlo abrió la caja de Pandora, y del mismo modo que Pandora liberó al hacerlo todos los males conservando dentro solo la esperanza, los números acarrearon un sinfín de preguntas, algunas diabólicamente complicadas. Hace apenas 20 años que pudimos dar con la prueba del Teorema de Fermat. Pierre de Fermat (1601 – 1665) escribió en el margen de un libro *"…es imposible encontrar la forma de convertir cualquier potencia más alta que el cuadrado, en la suma de dos potencias de la misma clase…"* y mantuvo ocupada a la comunidad matemática hasta que Andrew Wiles dio con la prueba en 1995. Otras cuestiones básicas sobre las propiedades de los números aún siguen pendientes de ser dilucidadas. Por ejemplo, la conjetura de Golbach (Christian Golbach (1690 – 1764)) permanece aún abierta a pesar de la simplicidad de su enunciado: *"Todo número par mayor que 2 es suma de dos números primos"*, del mismo modo que 8 = 3+5. Lo mismo ocurre con la conjetura de Beal de la que los periódicos se hacían eco hace unos años pues el Sr. Andrew Beal, rico banquero tejano, ofrece por su resolución un millón de dólares. El problema que él mismo formuló en 1997 se resiste y el Sr. Beal se empieza a impacientar. El infinito matemático tuvo como misión cerrar la caja de Pandora de los números pues, no importa como de grande sea el número, siempre hay uno mayor. Solo el infinito puede superar y dominar a todos los números.

Hoy, tras los trabajos desarrollados en el siglo XIX para formalizar una teoría de conjuntos completa, que diese fundamento definitivo a las Matemáticas, sabemos que hay muchos infinitos, y que unos son más grandes que otros. Buzz Lightyear tenía razón: Hay siempre un más allá después del infinito, otro infinito más grande.

Ciencia, Mística y vida cotidiana se encuentran en el punto común del infinito que se representa con un símbolo que se parece a un ocho tumbado que se abraza a sí mismo, cubriéndolo todo, empleado por primera vez por John Wallis (1616-1703),

inspirándose en la forma de la curva “lemniscata” introducida por Jacob Bernoulli (1655-1705), del latín lemniscus, que significa “cinta colgante”.

El infinito es ubicuo, está en todas partes. Milan Kundera lo evocó de manera infinitamente simple y bella: *“Quien busque el infinito que cierre los ojos”*.

Enrique Zuazua
Doctor en Matemáticas
Director de la Cátedra de Matemática Computacional, DeustoTech, Universidad de Deusto, Bilbao
Catedrático en Matemática Aplicada – Universidad Autónoma de Madrid

Capítulo 4
¿Qué diferencia hay entre los números ordinarios y los números binarios y cuáles son las ventajas y limitaciones de cada uno?
(Por Pedro Alegría Ezquerra)

Uno de los mayores descubrimientos (¿o podemos llamarlo invento?) de nuestra cultura es el sistema de numeración decimal. La introducción en el mundo occidental del sistema posicional en base diez, que terminó por desbancar casi por completo al sistema de numeración romano, vino de la mano de los árabes, que a su vez lo importaron de la cultura hindú, quienes ya incluían el cero antes del siglo VII como una de las diez cifras que nos son tan familiares. Debemos agradecer a los matemáticos al-Jwarizmi (c. 780-850) y al-Kindi (c. 801-873) la difusión del sistema hindú en el Oriente Medio y en Occidente.

Por cierto, como el cero surgió para determinar una posición que no estuviera ocupada por ninguna cifra significativa, no era necesario para contar. Así que no ha habido año cero ni siglo cero, pero el edificio matemático se desestabilizaría si el cero tuviera esa labor posicional como único objetivo de su existencia.

La implantación en Europa del nuevo sistema hindú empezó en Italia gracias a los esfuerzos de Gerbert d'Aurillac (c. 946-1003), más tarde conocido como el papa Silvestre II, y se extendió en el resto a través de la Escuela de Toledo, durante el siglo XIII. Para una eficaz difusión, resultó muy importante la publicación del libro *"Liber Abaci"* (1202), de Leonardo Pisano (c. 1170-1250), donde explicaba con detalle el nuevo sistema.

Algunos sistemas de numeración posicional muy extendidos en otras épocas, como el de base 20 utilizado por los mayas, han desaparecido. Otros se mantienen de forma residual: en base 12 se cuentan todavía los huevos, las horas y los meses; en base 60 los minutos y los segundos. Por ejemplo, cualquier persona culta entiende que 3 centenas, 6 decenas y 5 unidades corresponden al número

$$365 = 3 \times 100 + 6 \times 10 + 5 = 3 \times 10 \times 10 + 6 \times 10 + 5$$

(y si no es así, conviene que empiece por leer el capítulo 5 de la precuela de esta obra *"100 preguntas básicas sobre ciencia*" de Isaac Asimov). Ahora bien, si nos referimos a 3 horas, 6 minutos y 5 segundos, debemos entender que el número total de segundos es

$$3 \times 60 \times 60 + 6 \times 60 + 5 = 11165.$$

¡Acabamos de convertir el número 365 desde el sistema sexagesimal (en base 60) hasta el sistema decimal (en base 10)!

Los sistemas de numeración en base 12 y 60 presentan algunas ventajas frente al sistema decimal: una docena de huevos puede empaquetarse en cartones rectangulares de 12 x 1, 6 x 2 o 4 x 3; sin embargo, una decena de huevos solo podríamos encontrarla en cartones de 10 x 1 o 5 x 2. En general, cuantos más divisores tenga la base de un sistema de numeración, más versátil es la factorización de los números. Por esa razón,

es muy cómodo el sistema sexagesimal: resulta que, como 60 = 2 x 2 x 3 x 5, hay muchas posibles factorizaciones de este número.

Como curiosidad, citaremos a la *Dozenal Society of America*, corporación sin fines de lucro fundada en 1944, organizada para

> *"conducir la investigación y educación pública en el uso de la base doce como numeración en cálculos matemáticos, en pesos y medidas, y otras ramas de la ciencia pura y aplicada."*

¿Por qué se ha impuesto en nuestra cultura el sistema decimal? No hay duda que nuestra forma de "contar con los dedos" ha sido la responsable de que el diez sea el elegido como base del sistema de numeración. Es probable que, si siguiéramos descalzos como los monos, estaríamos utilizando un sistema en base veinte. Esto conllevaría un problema significativo: necesitaríamos idear veinte símbolos distintos para representar las veinte cifras y nos costaría mucho aprender las correspondientes veinte tablas de multiplicar.

En el otro extremo está el sistema binario, el de base dos (reservado para extraterrestres que tuvieran dos dedos): solo tiene dos cifras, digamos 0 y 1. Las tablas de multiplicar son el sueño de todo estudiante:

$$0 \times 0 = 0;\ 0 \times 1 = 0;\ 1 \times 0 = 0;\ 1 \times 1 = 1.$$

Los números expresados en este sistema tienen un pequeño inconveniente: enseguida se hacen muy grandes. Por ejemplo, el número 365 en base dos se escribe como 101101101. Para comprobarlo, debemos hacer la operación

$$1 \times 2^8 + 0 \times 2^7 + 1 \times 2^6 + 1 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 = 365.$$

En la sociedad que nos rodea, el sistema binario se está imponiendo a marchas forzadas: todo lo relacionado con la informática y las telecomunicaciones, sean por cable o no, debe sufrir en algún momento un proceso de representación en base dos.

¿Por qué es así? Un ordenador no tiene dedos pero está compuesto por circuitos, pudiendo presentar cada uno de ellos uno de estos dos posibles estados: encendido (representado en este caso por el dígito 1) o apagado (representado por el dígito 0). Cualquier mensaje que queramos transmitir por medio de un ordenador, ya sea numérico o alfabético, debe primero traducirse a su lenguaje, es decir a una sucesión de dígitos binarios. En informática, un dígito binario recibe el nombre de *bit* (acrónimo de binary digit) y es la unidad de información más pequeña que procesa un ordenador. El nombre bit fue adoptado en 1948 por Claude Shannon (1916-2001), uno de los pioneros de la informática pero el mecanismo fundamental de funcionamiento de un ordenador fue ideado por el considerado padre de la computación Charles Babbage (1791-1871) en 1812, inspirado en el uso de tarjetas perforadas, por parte de Joseph Marie Jacquard, para realizar intrincados diseños en sus telares, lo cual supuso una revolución en la industria textil.

Esta nueva aritmética binaria tiene relación con otras disciplinas matemáticas: por ejemplo, el álgebra de proposiciones, ideada por Georges Boole (1815-1864), se basa en el conjunto de operaciones realizadas con los valores lógicos verdadero y falso, las cuales son equivalentes a las que se establecen con los dígitos 1 y 0.

El gran desarrollo de la informática actual se debe a la creciente capacidad de almacenamiento y al aumento de la velocidad de cálculo. ¿Esto significa que la historia

acaba aquí? Ni mucho menos: el matemático azerbaiyano Lofti Zadeh (n. 1921) introdujo en 1965 la llamada lógica difusa (o álgebra borrosa), en la que se admiten números comprendidos entre cero y uno. Del mismo modo que hay más colores que el blanco y el negro o que puede haber proposiciones que no son completamente verdaderas o completamente falsas, la lógica difusa permite tender un puente entre la lógica clásica y el mundo que nos rodea.

¿Quién puede aventurar el futuro de la computación cuando se dote a los ordenadores de la capacidad de hacer operaciones con esta nueva aritmética y, en consecuencia, de tomar decisiones en situaciones más próximas a la realidad?

Pedro Alegría Ezquerra
Doctor en Matemáticas
Profesor titular
Universidad del País Vasco/Euskal Herriko Unibertsitatea

Capítulo 5
¿Qué son los números imaginarios? ¿Tienen alguna aplicación en la vida cotidiana?

(Por Yves Huttel)

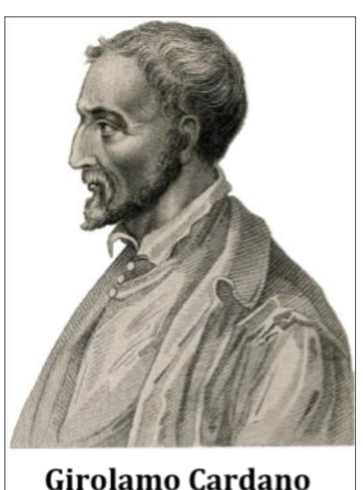
5.1 Introducción: ¿Por qué este capítulo?

Los descubrimientos humanos aparecen casualmente o como fruto de investigaciones. Me atrevería a decir que estos últimos surgen cuando las investigaciones son orientadas (típicamente investigaciones para resolver un problema concreto en cuyo caso se habla de investigaciones aplicadas) o básicas. Cuando las investigaciones son básicas, no tienen como propósito resolver un problema o aportar una solución concreta. En la sociedad actual se tiende a pensar que las investigaciones básicas tienen poco o ningún interés ya que "no sirven". Este capítulo tiene como objeto un caso concreto de investigación básica cuyos resultados han resultado ser de gran interés para múltiples/diversas aplicaciones. Como veremos, los números imaginarios no solamente son muy reales, sino que son muy importantes en nuestra vida diaria. A través de este ejemplo, ilustraremos la importancia de la investigación básica.

5.2 Definición y aspectos históricos.

Los números imaginarios forman parte de los números llamados complejos que se escriben de la siguiente forma: a + ib, siendo a y b números reales (un número real es un número que tiene una parte entera y una lista finita o infinita de decimales). En esta formulación, ib es lo que se llama la parte imaginaria del número complejo. La definición poco intuitiva pero genial de i es que $i^2 = -1$.

La primera aparición de un número imaginario data de 1545 en la forma de $\sqrt{-15}$ en un trabajo de Girolamo Cardano (también llamado Hieronymus Cardanus en latín o Jérôme Cardan en francés), matemático, filósofo, astrólogo, inventor y médico italiano (Pavia 1501, Roma 1576). Tanto Girolamo Cardano como otro matemático italiano, Raphaël Bombelli (Bolonia, 1526-1572) en su tratado de matemáticas "L'Algebra", mostraron el interés de utilizar las raíces cuadradas de números negativos en los cálculos matemáticos. Hasta el siglo XIX los números imaginarios eran considerados como un "truco" matemático imaginado (de allí su nombre) para, en particular, resolver ecuaciones del tercer grado i.e. del estilo $ax^3 + bx^2 + cx + d = 0$. Destaca la contribución en ese sentido de la escuela italiana, no solamente con Cardano y Bombelli, sino también de Niccolò Fontana (1499-1557) y Ludovico


Girolamo Cardano
(1501-1576)

Ferrari (1522-1565). Posteriormente varios sabios contribuyeron al desarrollo y uso de los números imaginarios a lo largo de los siglos como Gottfried Wilhelm Leibniz (1646-1716) o Leonhard Euler, matemático y físico suizo (1707-1783) que en 1777 define el número imaginario i tal que $i = \sqrt{-1}$. Curiosamente este mismo año, nace Carl Friedrich Gauss (1777-1855) quien también contribuirá ampliamente al uso de los números imaginarios.

Con el tiempo los números imaginarios se han "colado" en todas las áreas de física matemática, donde se utilizan los números complejos para resolver las ecuaciones, en magnetismo, electricidad, dinámica de fluidos, física cuántica, etc. De esta manera, los números imaginarios han dejado de ser una curiosidad matemática y han pasado a ser herramientas que permiten resolver problemas en áreas que indirectamente pero continuamente forman parte de nuestra vida como se ilustrará con algunos ejemplos a continuación.

5.3 Los números imaginarios bien reales en nuestras vidas.

Aunque aparentemente alejadas de nuestras vidas cotidianas, las matemáticas son parte integrante de nuestro día a día. Aunque no seamos del todo conscientes de ello, la investigación y evolución de los números imaginarios y más ampliamente de los números complejos en matemáticas han permitido resolver problemas matemáticos prácticamente irresolubles (por ejemplo el cálculo de algunas integrales). Estos avances en matemáticas han tenido y seguirán teniendo una repercusión en nuestras vidas. Pero independientemente de estos desarrollos puramente matemáticos los números complejos son herramientas muy importantes en diferentes campos como la ingeniería y la física que de forma más o menos directa afectan nuestras vidas. A continuación presentamos unos ejemplos (entre muchos otros) que ilustran aquello.

Los números complejos se utilizan para simplificar la modelización y la escritura de fenómenos oscilatorios como son las ondas electromagnéticas y los circuitos electrónicos. Si recordamos que una gran parte de las comunicaciones se realizan utilizando las ondas electromagnéticas (señal de televisión, radio, telefonía móvil…) y que los dispositivos electrónicos que utilizamos (ordenadores, teléfonos móviles, coches, etc.) contienen circuitos electrónicos, resulta evidente la presencia de los números imaginarios en nuestra vida diaria. Por otro lado se utilizan los números complejos en las series de Fourier que permiten el tratamiento y análisis de señales como son las señales electromagnéticas que, como hemos mencionado, se utilizan en particular en telefonía. Por lo tanto podemos decir que cuando utilizamos el teléfono móvil estamos utilizando de forma activa los números complejos y por lo tanto los números imaginarios.

Otro ejemplo es la mecánica de fluidos (hidrodinámica o aerodinámica) que estudia el comportamiento de los fluidos como puede ser el aire en los contornos de un avión o un coche. En mecánica de fluidos en 2 dimensiones (en un plano) se utilizan los números complejos que permiten una modelización más simple de los fenómenos como el flujo alrededor de un obstáculo. Una herramienta que utiliza los números complejos es la transformación conforme de Joukovsky que permite calcular el perfil de las alas de los aviones. Por lo tanto los números complejos están presentes en el diseño aerodinámico de coches y aviones y en el diseño hidrodinámico de barcos que a su vez permite una reducción de las fricciones y una reducción del consumo de carburante.

Menos intuitivo y directo es el ejemplo de la **mecánica cuántica**, para la cual la ecuación de Schrödinger (1925) es una ecuación fundamental que permite describir la evolución temporal de una partícula no relativista. Sin entrar en detalles, se observa claramente que el número imaginario "i" aparece en la ecuación. De la misma forma, los números complejos son utilizados en las herramientas de la mecánica cuántica (espacio complejo de Hilbert, matriz de Heisenberg). En resumen los números complejos se utilizan para explicar el comportamiento de la materia a nivel cuántico. Esto significa que la mecánica cuántica permite describir los fenónemos a nivel atómico como la dualidad onda-corpúsculo y la computación cuántica que rige los ordenadores cuánticos (1).

$$H(t)|\psi(t)\rangle = i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle$$

Como último ejemplo me gustaría comentar sobre la utilización de los números complejos en el estudio de un tema con el que no nos topamos todos lo días pero que llevamos en nosotros: ¡el origen del universo! Según la teoría del Big Bang (o gran explosión) el universo estaba en un estado muy condensado y luego se expandió (con una gran explosión). En ese modelo, si se extrapolan las leyes de la física hacia el origen del universo, nos encontramos con una singularidad (un punto) que estaría aproximadamente a 13800 millones de años (que sería la edad del universo). Stephen Hawking y James Hartle han postulado la hipótesis de un universo sin bordes donde la singularidad inicial no existiría. Esta hipótesis está basada en la idea de que el tiempo "τ" cerca del origen del universo es un tiempo imaginario que se define como $\tau = i\ t$. Según Hawking y Hartle esta formulación del tiempo permitiría describir la física del universo cerca de sus orígenes (cerca del Big Bang).

5.4 Conclusiones.

Hemos visto que, lo que inicialmente se considero como un "truco" matemático para resolver ecuaciones en el siglo XVI, y que se llegó a llamar "imaginario" por su extravagancia, está siendo utilizado en nuestros días para resolver problemas de nuestra vida cotidiana. A través de este ejemplo se evidencia la importancia de la investigación básica que por muy "inútil" que parezca puede tener aplicaciones e implicaciones muy importantes en el futuro. En una sociedad donde todo debe ser útil a corto plazo, no cabria la posibilidad de financiar la investigación básica que diera lugar a números "imaginarios" ya que todo tiene que ser real. El caso de los números imaginarios no es único y existen otros ejemplos de resultados de estudios e investigaciones básicas que dieron lugar a importantes aplicaciones. De allí la necesidad de preservar la investigación "imaginaria" para el avance y bienestar de nuestra sociedad.

Notas:
1: "Cómo los ordenadores cuánticos cambiarán para siempre la computación"

Yves Huttel
Doctor en Física
Científico Titular del CSIC

Capítulo 6
¿Qué es la topología?
(Por Marta Macho Stadler)

De manera informal, la topología es la parte de las matemáticas que estudia propiedades cualitativas de espacios y objetos. La topología se ocupa de aquellas características de las figuras que permanecen invariantes cuando éstas son plegadas, dilatadas, contraídas o deformadas, de modo que no aparezcan nuevos puntos –no se pueden *romper* los objetos estudiados– o se hagan coincidir puntos diferentes –no se pueden *pegar* puntos que no lo estuvieran previamente–. Por ejemplo, en topología, un balón de rugby, un balón de fútbol o una pelota de ping-pong son *indistinguibles* porque, si estuvieran fabricados de un material maleable –por ejemplo plastilina–, sería posible deformar los unos en los otros sin romper ni pegar nada. Por decirlo de una manera más transparente, en topología no son importantes ni las posiciones de los objetos, ni los tamaños, ni las formas: los tres ejemplos que hemos comentado antes son *superficies* que encierran volumen vacío. Esa es la cualidad topológica que los define.

Las transformaciones permitidas en topología presuponen que hay una correspondencia biunívoca entre los puntos de la figura original y los de la transformada, y que durante una de estas deformaciones se hacen corresponder puntos próximos a puntos próximos. Esta última propiedad se llama *continuidad*, y lo que se exige es que la transformación y su inversa sean ambas continuas: trabajamos con lo que llamamos *homeomorfismos.*

Los objetos de la topología son los mismos que los de la geometría, pero se trabaja con ellos de manera diferente: un círculo es equivalente a una elipse; una bola no se distingue de un cubo: se dice que la bola y el cubo son objetos topológicamente equivalentes –*homeomorfos*–, porque se pasa, como ya hemos indicado, de una al otro mediante una transformación continua y reversible.

Uno de los objetivos fundamentales de la topología es distinguir espacios. Si sabemos que dos objetos son homeomorfos y queremos demostrarlo, debemos encontrar una transformación –una aplicación continua– que deforme el uno en el otro y que además sea reversible –también de manera continua–. Aunque puede que resulte algo complicado, sabemos lo que tenemos que hacer. Sin embargo, cuando queremos demostrar que dos objetos no son homeomorfos, a priori habría que ver que no hay ninguna de esas especiales transformaciones reversibles que lleve el uno en el otro. ¿Y eso cómo se hace? ¿Debemos ir probando con todas las transformaciones que se nos ocurran? ¿Y cómo estamos seguros de que no se nos olvida ninguna? Para poder distinguir objetos topológicamente diferentes, se introducen los denominados *invariantes topológicos*, es decir, nociones que permanecen inalterables al cambiar un objeto por otro topológicamente equivalente. Por ejemplo, algunos de ellos son la compacidad, la conexión, la propiedad de Hausdorff, el tipo de homotopía, etc. Por

ejemplo, si un espacio es conexo –intuitivamente, 'de una pieza'– y otro no lo es, se puede concluir que no son homeomorfos. ¡Cuidado! Eso no significa que dos espacios conexos sean siempre topológicamente equivalentes: por ejemplo, una bola maciza y una pelota son conexas y no son homeomorfas –es imposible 'eliminar' el agujero que encierra la pelota–.

Aunque en muchas ocasiones los desarrollos teóricos en topología –o en otras ramas de las matemáticas o de la ciencia– no han tenido o tienen una aplicación inmediata, el investigar en esos campos puede, de manera indirecta, ayudar en el avance de otras disciplinas. Por este motivo, los equipos de investigación son cada vez más interdisciplinares: el combinar conocimientos de áreas diversas, con *miradas* y formaciones diferentes solo puede contribuir a mejorar cualquier estudio.

Un ejemplo sorprendente de la utilidad de la topología es el de la llamada *teoría topológica de nudos.* Los nudos están presentes en ámbitos tan dispares como la decoración, la industria textil, la magia, el alpinismo o la cirugía. Su estudio matemático –la *teoría topológica de nudos*– ha permitido descubrir su relación con la física, la química o la biología molecular.

En matemáticas, un nudo se piensa como una curva continua, cerrada y sin puntos dobles situada en un espacio tridimensional. Dos nudos son *equivalentes* cuando es posible pasar de uno a otro mediante deformaciones, estiramientos o compresiones, pero sin realizar cortes. Es muy difícil decidir cuando dos nudos son equivalentes, y gran parte de la teoría de nudos se dedica precisamente a intentar resolver esa cuestión. Algunos trucos de magia utilizan justamente esta propiedad: el ilusionista nos presenta una cuerda *anudada* de manera complicada y usando su destreza –y algunas tretas añadidas para despistar– deshace ante nuestros ojos las ataduras sacudiendo con fuerza la cuerda. En realidad, el mago ha partido del *nudo trivial* –no hay nudo– presentado de una manera complicada para disimular la realidad de esa atadura. No ha hecho magia, ha hecho topología.

El ADN posee una estructura de doble hélice en la que dos cadenas de nucleótidos complementarios se enrollan a lo largo de un eje común. Esta doble hélice puede moverse en el espacio para formar una nueva hélice de orden mayor: se habla en este caso de ADN superenrollado. El ADN circular superenrollado es una doble hélice de moléculas donde cada cadena de polinucleótidos forma un anillo. Todas las propiedades físicas, químicas y biológicas del ADN están influenciadas por la circularidad y las deformaciones asociadas al superenrollamiento. Es posible comprender este mecanismo de superenrollamiento –y las consecuencias de esta estructura para el ADN– utilizando matemáticas complejas, en particular topología. Para realizar este estudio se comienza construyendo un modelo matemático representando la estructura helicoidal del ADN y –entre otros factores– es necesario describir los nudos que aparecen en la configuración, encontrar las características esenciales que permitan distinguirlos, es decir, clasificarlos sin riesgo a confusión. Estas características, que deben permanecer inalterables a lo largo de la deformación son los *invariantes topológicos* del nudo. Las *topo*isomerasas son enzimas capaces de actuar sobre la topología del ADN: lo enredan o desenredan –es decir, deshacen o crean nudos– para permitir un almacenamiento más compacto o facilitar su replicación. La comprensión del funcionamiento de estas enzimas y su interacción con el ADN podría ayudar a conocer mejor algunas enfermedades genéticas. En esta tarea, la topología tiene mucho que decir.

La topología se utiliza en muchas más ramas de la ciencia: en el estudio de flujos –como la atmósfera alrededor de nuestro planeta–, en el análisis de redes de diversos tipos, en cosmología –como el examen de la forma del universo–, en física de materiales –como en el estudio de cristales líquidos–, etc. Es realmente emocionante ver cómo una teoría procedente de la matemática pura encuentra aplicaciones en ramas tan diversas de la ciencia… ¡y las que aún estarán por llegar!

Marta Macho Stadler
Doctora en Matemáticas
Profesora Facultad de Ciencia y Tecnología,
Universidad del País Vasco-Euskal Herriko Unibertsitatea

Capítulo 7
¿Qué es el azar?
(Por Lance Fortnow)

Todos los días lidiamos con el azar. El hombre del tiempo dice que habrá un 30% de posibilidades de que llueva hoy. Lanzamos una moneda al aire para decidir qué película vamos a ir a ver al cine. El precio de la póliza de nuestro coche se calcula en base a la probabilidad que la compañía aseguradora cree de que tengamos un accidente.

Así, llueve hoy, la moneda sale cara, y no tenemos ningún accidente. ¿Hay alguien que elija estos resultados o están predeterminados? Y si así lo fuera, ¿por qué pensamos que se deben al azar? Este artículo no trata sobre la probabilidad, pero sí sobre cómo el azar ocurre, o parece ocurrir, en nuestras vidas cotidianas.

7.1 Lanzando monedas al aire.

Miremos al proceso de lanzar una moneda al aire. Nuestro pulgar golpea la moneda y hace que ésta gire sobre sí misma una y otra vez. La fuerza del pulgar, la trayectoria de la moneda y también, por qué no, la presión y la resistencia del aire controlan el giro. Cuando la moneda golpea el suelo, dependiendo de de su ángulo, ésta caerá en uno de los dos posibles estados de baja energía, mostrando la cara o la cruz.

Particularmente, no hay nada de azar en este proceso. Cualquier variable podría ser controlada y simulada. El que salga cara o cruz está determinado en el momento en el que la moneda se aleja del pulgar. Aún así, en los partidos de fútbol se les pide a los jugadores que elijan un lado de la moneda cuando ésta está aún en el aire, tratando al proceso como si fuera aleatorio.

El clima meteorológico y la seguridad que tenemos al conducir dependen de cadenas de eventos mucho más complejas, pero aún así siguen determinadas por las condiciones iniciales. Así, aparecen dos preguntas que merece la pena plantearse:

¿Por qué consideramos estos procesos como aleatorios?

¿Existe verdaderamente la aleatoriedad en la Naturaleza?

Dejadme responder a ésta última primero:

7.2 El azar en la Naturaleza

“Dios no juega a los dados” así aclamaba Albert Einstein su creencia en el determinismo científico. Antes del siglo XX muchos científicos pensaban de igual manera, que el mundo y el universo en general se mueven siguiendo una trayectoria

totalmente definida por su estado actual. Esta visión fue cuestionada durante el siglo XX, y gracias en parte al desarrollo de la mecánica cuántica.

Tomando un simple ejemplo, supongamos que tenemos una bombilla y ponemos justo al lado un trozo de cartón al que le hemos hecho una diminuta ranura vertical. Cuando la bombilla se enciende, la luz que va a pasar a su través va a estar orientada en la dirección vertical. Esta orientación se puede medir fácilmente poniendo un segundo trozo de cartón con otra hendidura a continuación. Si lo orientamos de igual forma que el primero, esto es con la ranura en vertical, veremos que pasa luz a su través; en cambio, si lo orientamos de manera horizontal no habrá ni rastro de luz.

¿Pero qué pasaría si orientamos la segunda ranura formando un ángulo de 45 grados con la primera? Entonces solo la mitad de la luz pasará por ambas. Si disminuimos la luz que emite la bombilla también reduciremos la cantidad de luz que pasa por la segunda rendija a 45 grados.

Según la mecánica cuántica la luz no está hecha de una sustancia a la que podamos reducir su tamaño tanto como queramos. No, la luz está compuesta por pequeños paquetes llamados cuantos o fotones que componen un haz, al igual que los granos de arena componen una playa. Si en nuestro ejemplo de la bombilla reducimos la luz emitida a un simple cuanto de luz, ¿qué pasaría si un fotón verticalmente orientado golpea la ranura a 45 grados?

Esta pregunta es fácil de responder si colocamos un foto-detector al otro lado del segundo cartón, que accione una campana cada vez que un fotón pase a su través. Sin sorpresa veremos que en la mitad de las ocasiones el fotón será bloqueado por la ranura a 45 grados, y que en la otra el fotón pasará haciendo sonar la campana.

Esto sí que parece verdadero azar, un experimento totalmente controlado y reproducible que tiene dos posibles resultados: que una campana suene o no, teniendo una probabilidad del cincuenta por ciento ambas. Verdaderamente Dios parece jugar a los dados para decidir si la campana suena o no.

¿O igual no? Quizá simplemente estamos observando un trozo de un sistema determinista aún más grande, y midiendo el fotón reducimos la dimensión del sistema que observamos, pero aún así sigue formando parte de uno aún mayor. Esto parece confuso de primeras, así que dejadme suponer qué es lo que ocurría si no pudiésemos observar el resultado del experimento. Supongamos que en vez de hacer sonar una campana, el detector libera un gas venenoso dentro de una caja en la que hay un gato. Si el fotón es detectado, el gato es envenenado; si no, el gato continua viviendo sin darse cuenta del cruel experimento que está ocurriendo a su alrededor. Si suponemos también que no miramos dentro de la caja, acabamos de construir el famoso experimento de Schrödinger.

Si no miramos dentro de la caja, es obvio que no podemos saber si el gato está con vida o no. Podemos pensar que está vivo o muerto, pero realmente no sabemos la respuesta. En tal tipo de situaciones podemos definir lo que se llama un estado cuántico, donde el gato puede estar vivo y muerto a la vez en una llamada superposición cuántica de estados. Es cuando abrimos la caja cuando la superposición colapsa en uno de los dos estados, muerto o vivo. De igual manera, una persona situada fuera de nuestro universo

podría describirnos como una serie determinista de estados cuánticos en superposición, siempre y cuando nunca mirase dentro del sistema.

Entonces, ¿el verdadero azar es debido a la naturaleza mecano-cuántica de la realidad o no? No hay respuesta clara para esta pregunta. De hecho, la pregunta plantea más un debate filosófico que científico. Además hay otras fuentes potenciales de azar como son los agujeros negros, que parecen destruir información y de los cuales es incluso más difícil saber por qué lo hacen.

En definitiva los físicos parecen no tener una idea clara de si en la Naturaleza tenemos verdadera aleatoriedad o no. Sin embargo, esto nos aleja del experimento de la moneda al aire. Como dijimos, si el lanzamiento se realiza en las mismas (exactas) condiciones la moneda acabará siempre del mismo lado. Entonces, ¿cómo ocurre el azar en procesos potencialmente deterministas?

7.3 Azar y complejidad

Si el lanzar una moneda al aire es un proceso totalmente determinista entonces, ¿por qué consideramos el que salga cara o cruz como un resultado aleatorio? Como dijimos, si seguimos el movimiento de la moneda en el aire, usando los sensores adecuados y el poder de cálculo suficiente, podríamos determinar el resultado de la moneda con total seguridad. Sin embargo, en las situaciones cotidianas en las que lanzamos una moneda al aire es obvio no tenemos al alcance ese poder de computación, y por ello asumimos que el resultado es completamente impredecible. Esta impredecibilidad de la respuesta nos hace tratar al evento como si fuera completamente azaroso, aunque sea verdaderamente nuestra incapacidad para calcular el resultado lo que le da esta cualidad.

Lo mismo se puede aplicar a la predicción del tiempo. Las agencias meteorológicas usan herramientas y computadoras muy potentes para predecir el tiempo, pero sus predicciones están basadas en modelos muy limitados ya que incluso la máquina más poderosa que podamos pensar, es incapaz de tomar en cuenta todos los factores necesarios para predecir incluso el tiempo que hará mañana. Aún así, el hombre del tiempo suele dar el porcentaje de la probabilidad de que mañana llueva, tratando la lluvia como un elemento aleatorio.

Pongamos el ejemplo de un casino. La ruleta es un dispositivo simple pero su resultado está basado en interacciones tan complejas que parece que sea un proceso completamente aleatorio. De hecho, aprovechando esta complejidad los casinos literalmente invierten dinero asumiendo que los jugadores no pueden más que acertar por azar el resultado de una tirada de dados o el giro de la ruleta. Muchos casinos permiten incluso a los jugadores lanzar los dados, sabiendo que incluso haciendo esto los jugadores no tienen ninguna ventaja significativa en averiguar el resultado de antemano. En el juego del blackjack el croupier baraja las cartas delante de los jugadores con la certeza de que la nueva disposición de las cartas será completamente aleatoria. La gente experta dedicada al conteo de cartas toma esta nueva disposición de la baraja como si fuera completamente al azar. Esto les ayuda a predecir qué cartas van a salir en la próxima mano usando la información de las anteriores.

En los mercados financieros, los agentes de bolsa ponen precio a los valores bursátiles asumiendo la probabilidad de que un producto valga un determinado valor en el futuro. Esto es tan solo una aproximación ya que los precios reales de mercado dependen de una forma muy compleja de lo que vaya a suceder en la compra venta durante ese tiempo.

En Gran Bretaña, el apostar está tan difundido que no solo puedes hacerlo sobre resultados deportivos sino también sobre el resultado de unas elecciones o incluso sobre quién será el ganador en una determinada gala de premios. Los sitios de apuestas no corren ningún riesgo ya que solo plantean apuestas donde las probabilidades de que salga un resultado u otro sean similares. Así hacen dinero independientemente del resultado final. Sin embargo, nosotros como jugadores hacemos estimaciones de probabilidad (consciente o inconscientemente) sobre a cuál de los resultados merece la pena apostar.

Incluso dejadme considerar el ajedrez. En este juego se puede decir que no hay azar en absoluto. Las posiciones de las fichas en el tablero están a la vista de ambos jugadores, además no hay ningún elemento de aleatoriedad como los dados en el backgammon o el baraje de cartas en el poker. Sin embargo, se sigue hablando de probabilidades, de cuán probable es que las fichas blancas ganen a las negras después de efectuar un determinado movimiento. Se dice que la complejidad del ajedrez traslada el juego de “perfecta información” a un juego de “imperfecta información”, añadiendo una medida de azar a un juego que no tiene ninguna fuente concreta de él.

Cuando preguntamos a una computadora que nos genere un número aleatorio, realmente no lo hace. Lo que nos proporciona es en realidad el resultado de un cálculo muy complejo, tanto que lo podemos considerar como aleatorio. De igual manera, si leemos un mensaje al que se le ha aplicado un protocolo de cifrado y del cual no tenemos la llave para decodificarlo, nos parecerá que ha sido generado completamente al azar. Estudios teóricos han mostrado cómo cualquier función compleja puede ser traducida mediante generadores de pseudo-azar y protocolos criptográficos, a un resultado que es imposible distinguir de la verdadera aleatoriedad. De forma práctica se han llegado a desarrollar protocolos de cifrado tales que ni hombre ni máquina son capaces de distinguir del puro azar.

7.4 Disminuyendo el azar batallando la complejidad

Con el desarrollo de algoritmos más eficientes y computadoras más potentes, hoy en día ya somos capaces de hacer frente al azar proveniente de la complejidad de la Naturaleza. Las nuevas tecnologías han permitido acceder a cantidades masivas de información que, combinada con métodos de aprendizaje máquina (machine learning), nos ayuda a vencer la barrera de la incertidumbre. Aunque usando el cálculo de estas probabilidades seguimos sin poder predecir el futuro con una absoluta certeza, si que podemos obtener información valiosa que sin duda nos dan una ventaja considerable frente aquellos que simplemente miran los sucesos como si fueran completamente impredecibles.

Nuevos modelos teóricos, computadoras más potentes y mejores algoritmos han mejorado considerablemente la predicción del tiempo meteorológico, aunque aún estemos lejos de predecirlo con absoluta seguridad. Los fondos de cobertura usan

técnicas matemáticas para ganar ventaja en la compra venta de valores. Usando dispositivos electrónicos especiales, los apostadores profesionales son capaces de encontrar pequeñas imperfecciones en las ruedas de las ruletas, tales que esta información les proporciona ventajas en sus apuestas. Hoy en día cualquier juego de ajedrez electrónico, incluso aquél de nuestro smart phone, es capaz de ganar a cualquier ser humano. La capacidad de la máquina para estimar la probabilidad de ganar en cada uno de los posibles movimientos, es mucho mejor de la que ningún ser humano nunca podría llegar a tener.

7.5 ¿Qué es el azar?

La pregunta sobre si en la Naturaleza obtenemos verdadero azar o no, depende de lo que interpretemos de lo que en la Naturaleza realmente ocurre. De hecho, lo que definimos como azar, no lo es en absoluto: es simplemente la consecuencia de nuestra incapacidad para poder predecir el resultado de un proceso complejo.

El desarrollo de nuevas herramientas de análisis de datos y “machine learning” nos ayuda a realizar predicciones más fiables. Sin embargo, seguirán existiendo procesos tan complejos que de los cuales nunca seremos capaces de predecir su resultado. Lo mejor que podemos hacer es entender la verdadera naturaleza del azar. La toma de decisiones frente a la incertidumbre es uno de los retos que todos afrontamos en el día a día. Incluso grandes líderes políticos y económicos toman decisiones que más tarde lamentan al ver como los hechos acontecen. Aún así, entendiendo lo que no podemos predecir nos da mejores herramientas para afrontar los desafíos del futuro.

Se agradece la ayuda del doctor **Pablo López Tarifa** por una traducción precisa de este capítulo.

Lance Fortnow
Doctor Matemática Aplicada
Professor and Chair
School of Computer Science, Georgia Institute of Technolgy

Capítulo 8
¿El universo tiene borde?
(Por Elena Denia)

There was a young lady named Bright,
Whose speed was far faster than light;
She started one day
In a relative way,
And returned on the previous night.

Arthur Henry Reginald Buller (1874 – 1944)

Todos alguna vez hemos escrutado la línea del horizonte, aguzando la vista para encontrar un atisbo de tierra o el final de los océanos. A lo largo de la historia, desde el más tenaz de los conquistadores hasta el individuo que da un paseo desinteresado por la costa, se ha preguntado por los bordes de la gran masa de agua. «Más allá hay dragones», decían unos. «Es un abismo sin límite», aseguraban otros. Las playas siempre han sido un escenario evocador.

Con la misma naturalidad hemos levantado la mirada hacia los cielos. Si bien es cierto que lidiar con la incertidumbre no es tarea fácil, el misterio siempre ha tenido un atractivo intrínseco al ser humano; curiosidad, afán de conocimiento –si se prefiere–; muchos de nosotros nos hemos preguntado alguna vez: ¿el universo tiene borde?

Al pensar en los confines del cosmos uno se puede imaginar en su busca, avanzando incesantemente en una nave a través del espacio. Este experimento mental permite plantear la idea de que, en algún momento, se acabarán los planetas, las estrellas, el polvo y todo lo demás. Que nos enfrentaremos a un vacío oscuro que aguarda alguna extraña deformación del tejido espaciotemporal, donde todo se acabe. A priori parece razonable, sobre todo después de escuchar a los expertos decir que lo más probable es que el universo sea plano.

Estamos ante un error conceptual que numerosas veces hace desistir al interesado. A menudo se dice que el universo es plano (geometría euclídea), esférico o con la forma de una silla de montar a caballo (geometría hiperbólica). Estas ideas generan confusión, porque uno se imagina el cosmos, por ejemplo, como la base de una pizza colosal o con la forma de una gran sandía. Sin embargo, cuando se citan esas tres posibilidades se está haciendo referencia a la geometría del tejido del universo en un entorno local, no a su morfología como un objeto único.

En el escenario local, una ciudad puede considerarse plana. Pero si pensamos en todo un continente, la curvatura de la Tierra empieza a manifestarse. En la ciudad, la suma de los ángulos de un triángulo siempre será 180º. En cambio, si trazamos el triángulo sobre el globo terrestre, la suma de sus ángulos será superior. Y si lo trazamos después sobre una silla de montar, veremos que es inferior.

Algo parecido sucede con el tejido del cosmos. No se trata de lo inmenso que sea el dibujo de nuestro triángulo, sino de comprender que el espaciotiempo es un tejido que se deforma con la presencia de masa, y la curvatura ya no se produce sobre la superficie terrestre, sino en tres dimensiones espaciales y una temporal, algo un poco más complejo de imaginar, pero todavía intuitivo. En cualquier caso este tejido,

mayoritariamente, presenta una geometría cercana a la euclídea. Esto significa que los triángulos que pintemos con una brocha sideral sumarán típicamente 180º, salvando algunas pequeñas deformaciones ocasionadas por la materia –de hecho, si uno observa con cuidado, descubrirá que nuestra ciudad también presenta algún que otro desnivel–.

Además, pensando a grandes escalas, el universo es homogéneo e isótropo –éste es el llamado principio cosmológico–. Homogéneo porque al comparar dos fragmentos cualesquiera, eso sí, lo suficientemente grandes, presentan un aspecto similar; con cúmulos de galaxias y filamentos que conectan tales estructuras. E isótropo porque podemos mirar en direcciones arbitrarias y también presentará un aspecto similar.

Para ilustrarlo de forma más mundana, solo hay que pensar en que desde cualquier posición en la superficie de la Tierra podemos mirar hacia la profundidad del cosmos, sin que encontremos un lugar preferido para capturar la luz –excepto, por supuesto, aquellos espacios más despejados donde situar los telescopios–. El universo se nos presentará, aproximadamente, con el mismo aspecto en todas direcciones. Se mire como se mire, ahí estarán los cúmulos y los súper cúmulos; ya sean de polvo, de estrellas o de galaxias.

Y nosotros los observamos porque atrapamos sus partículas de luz –*fotones*– con los telescopios. Y es de esa luz de donde extraemos la información más variada, desde la composición química de un astro concreto hasta el alejamiento acelerado de una galaxia respecto a la nuestra.

De hecho, mirar hacia el universo profundo se traduce en capturar la luz más antigua que nos llega: el fondo cósmico de microondas, la radiación que se liberó hace cerca de 14 mil millones de años, poco después de la gran explosión que marcó el inicio de los tiempos, el Big Bang.



Figura 1: *El universo a escala logarítmica. Imagen creada por el artista Pablo Carlos Budassi como regalo de cumpleaños para su hijo, a partir de imágenes de la Nasa y mapas de la Universidad de Princeton.*

Así, desde cualquier punto del globo terrestre, tendremos un límite visible al mirar hacia el espacio, un límite que en términos de distancia alcanza esos 14 mil millones de años luz. Pero siempre será el mismo, observemos desde España o desde las Islas Seychelles. Esta distancia define el radio del universo observable, que por extensión supone un *horizonte observable* que encierra la porción del cosmos que somos capaces de apreciar.

Hasta aquí tenemos frente a nosotros el borde del *universo observable*, situado a unos 14 mil millones de años luz apuntando en todas direcciones. Sin embargo, no podemos ignorar el hecho de que estamos mirando constantemente hacia el pasado, pues los fotones ancestrales que nos llegan provienen desde casi el principio de los tiempos. Quizá este borde no satisfaga al lector, al fin y al cabo, ¿podemos sentirnos aún contenidos en la piel del bebé que fuimos?

Para no perder la cordura, parece razonable retomar nuestro experimento mental y volver a surcar el espacio en nuestra intrépida expedición. Comprobaremos, una vez en marcha, que nuestro reloj de bolsillo no dejará de hacer «tic tac»: la flecha del tiempo siempre apunta hacia adelante. De este modo escaparemos del sistema solar, atravesaremos otros sistemas planetarios, también los confines de nuestra galaxia y los cúmulos globulares que la orbitan. Y todos estos objetos continuarán envejeciendo a nuestro alrededor, cada uno a su propio ritmo, con el devenir de un «tic tac» diferente –pero esto ya es otra historia–. ¿Nos encontraremos, entonces, con algún tope si viajamos a través del espacio indefinidamente?

Para tropezar con esa frontera debemos superar la velocidad de la luz y así poder abarcar una visión más allá del *universo observable*, porque precisamente esa es nuestra limitación: no podemos acceder a las partes del universo cuya distancia es tan prolongada que la luz no ha tenido tiempo suficiente para llegar hasta nosotros. Dadas tales complicaciones, no habrá más remedio que abortar nuestra misión por cuestiones técnicas. No obstante, al margen de esa restricción física, ¿quién nos impide tratar de deducirlo de manera teórica? Al menos –de momento– se pueden garabatear algunas soluciones interesantes sobre el papel e imaginarlas de forma conceptual en nuestras mentes.

Pensemos entonces, en la recta final de este ejercicio abstracto, en la geometría global del universo, que abarca tanto el fragmento observable así como todo lo demás.

Tratamos de dibujar el espacio métrico que ocupa el cosmos, la forma de su recipiente. Es aquí cuando debemos hacer el mayor esfuerzo y empujar nuestras mentes hacia el cambio de paradigma que supuso la relatividad general. Porque ya no podemos considerar al universo como un ente estático que podamos medir. Porque el tiempo transcurre. Y lo hace de forma distinta en las diversas partes. La cuestión, por tanto, no solo concierne al *cómo es*, sino también al *cuándo lo ha sido*, y por ello nuestro empeño por congelarlo para poder sacar la escuadra y el cartabón no es más que pura testarudez. Obstinación. Resistencia a entender que el tiempo fluye a distintas velocidades y que la sustancia espaciotemporal es solo una. No hay tiempo sin el espacio que le corresponde, ni espacio para el que no le es propio un tiempo.

Pese a todo ello, y para no defraudar al lector, puede quedarse con la idea de que, en realidad, no sabemos si el universo se extiende indefinidamente o si está contenido en un espacio métrico limitado. En el primer caso, tendríamos infinitos

caminos que seguir con nuestra nave, como en el plano euclidiano. En el segundo caso, el confinado o *compacto*, podríamos descubrir, con gran sorpresa, que siguiendo un camino llegásemos al punto de partida, como sucede con las geodésicas de una esfera. «No se sabe la forma *absoluta* del universo», podría decir tranquilo, sin verse demasiado alejado de su zona de confort y acariciando en su bolsillo el reloj familiar; con el capricho de que el «tic tac» sea unívoco en cada recoveco cósmico.

Sin embargo, lo inquietante de todo este asunto es que preguntarse por la forma del universo, muy posiblemente, sea un sinsentido. Porque la forma, ¿la forma, *cuándo*?

Elena Denia
MSc en Física Avanzada, especialidad Cosmología
Investigadora del Consejo Superior de Investigaciones Científicas (CSIC)
Instituto de gestión de la innovación INGENIO
Universidad Politécnica de Valencia

Capítulo 9
¿De qué está hecho el espacio?
(Por José Luis Fernández Barbón)

¿Qué es el espacio vacío? Sin duda, una pregunta de fundamentos que ha sido objeto de atención más filosófica que científica a lo largo de los siglos. Lo cierto es que, en los albores del siglo XXI, no sabemos realmente qué es el espacio, pero sí tenemos una gran cantidad de información acerca de su estructura. Una información destilada con cuentagotas a lo largo de tres milenios.

Probablemente, el paso más importante en esta epopeya fue el primero, cuando los griegos inventaron la geometría, su gran contribución a la historia del conocimiento. La geometría griega fue la primera gran teoría científica de la historia, la primera que podemos considerar "correcta", en el sentido de que nunca será suplantada en su dominio de aplicación. Todos esos hechos familiares acerca de triángulos y círculos que aprendemos en la escuela primaria caracterizan lo que podríamos considerar la "estructura" del espacio.

El siguiente paso cualitativo tuvo lugar en el siglo XIX, cuando los matemáticos se dieron cuenta de que la geometría euclidiana no era lógicamente necesaria. De pronto, era posible imaginar que, al medir con mucha precisión los tres ángulos de un triángulo, la suma resultara ligeramente diferente de 180^0. La caracterización final de todas las posibles desviaciones de la geometría griega la dio Riemann, que elucidó en toda generalidad el concepto de curvatura. Así, a finales del siglo XIX la geometría del espacio vacío empieza a considerarse, después de todo, una cuestión experimental.

Al mismo tiempo, la física entra en una crisis de fundamentos en relación con el comportamiento de la luz, una crisis de la que emergerán nuestras concepciones actuales del espacio, el tiempo, y la materia. Einstein es la figura dominante en este período revolucionario. En el "año milagroso" de 1905 sentó las dos bases sobre las que descansa el edificio de la física moderna: los conceptos de espacio-tiempo y de partícula cuántica. En un sentido muy literal, toda la física del siglo XX es una elaboración de estas dos ideas fundamentales: la energía se organiza en "paquetes", que llamamos partículas cuánticas, y que conforman tanto la "materia" como las "fuerzas". Todo esto sucede en un escenario geométrico llamado espacio-tiempo, una generalización de la geometría griega que incluye al tiempo.

El espacio-tiempo es una peculiar mezcla entre espacio y tiempo, análoga a la visión del espacio ordinario como una mezcla entre las diferentes direcciones. Estamos familiarizados con la idea de que el noroeste es una mezcla de las direcciones norte y oeste. Para cambiar de dirección en el espacio, efectuamos un giro por un cierto ángulo. Análogamente, para cambiar de dirección en el espacio-tiempo, efectuamos un movimiento a una cierta velocidad. Diferentes observadores llaman "espacio" a

diferentes "secciones transversales" del mismo espacio-tiempo aunque, al igual que ocurre con las diferentes direcciones del espacio, todas estas posibilidades son equivalentes entre sí.

Existen también algunas diferencias entre la descomposición del espacio en direcciones y la descomposición del espacio-tiempo en espacio y tiempo separados. En el primer caso podemos mezclar cualesquiera direcciones porque todos los ángulos son posibles. En el caso de la mezcla entre espacio y tiempo, solo podemos efectuar movimientos a velocidades inferiores a la de la luz, que resulta ser la velocidad máxima en la naturaleza. Este requerimiento es equivalente a la imposibilidad de viajar al pasado, es decir, la imposición de una relación consistente de causalidad.

La culminación de la obra de Einstein fue el descubrimiento de que, en presencia de un campo gravitacional, las diferentes direcciones del espacio-tiempo dejan de ser equivalentes: el cociente entre el perímetro y el diámetro de una circunferencia, o la suma de los ángulos de un triángulo, depende de la posición de éstos con respecto al objeto que crea el campo gravitacional. Einstein describió esta situación diciendo que la gravitación es equivalente a la "curvatura" del espacio-tiempo. Literalmente, la energía contenida en el espacio produce una deformación de su estructura geométrica, como si fuera una especie de membrana elástica.

Al igual que ya hiciera Maxwell en el siglo XIX con el campo electromagnético, Einstein fue capaz de encontrar las ecuaciones que gobiernan estas deformaciones de curvatura, como función de la densidad y tipo de energía contenida en el espacio. Podemos resumir estas ecuaciones en algunas reglas intuitivas:

1) Si la energía está localizada, como ocurre con un conjunto de partículas elementales que se encuentran en el interior de una cierta región, la ley de Einstein nos dice que el espacio se "estira" en la dirección radial y se contrae en la dirección transversal, de tal forma que si fuéramos a dibujar un círculo alrededor de la materia localizada, el cociente entre el perímetro y el diámetro sería menor que π. Toda energía localizada se puede analizar como un conjunto de partículas elementales de diferentes tipos, pero todas ellas tienen energía positiva con respecto al vacío.

2) Si hay energía en el propio vacío, el efecto solo es visible a grandes distancias. En el caso de que la energía del vacío sea negativa, el espacio se curva de forma que funciona como una caja de paredes suaves: tiramos una piedra y vuelve como un bumerán. Por el contrario, si la energía del vacío es positiva, el espacio se expande exponencialmente: cada cierto tiempo fijo, el volumen de espacio se dobla, produciendo una separación exponencial de cualesquiera partículas existentes en su interior. A la energía del vacío se le ha llamado "constante cosmológica" y más recientemente, "energía oscura".

A lo largo del siglo XX, los físicos han corroborado la ley de Einstein en todas las situaciones en las que el test cuantitativo ha sido posible. La última, y seguramente la más espectacular, es la detección directa de las ondas gravitacionales. Si el espacio-tiempo es "elástico", debería transportar ondas de curvatura cuando una distribución de energía cambia con el tiempo. Según la teoría de Einstein, el "coeficiente de elasticidad" del espacio es la constante de gravitación de Newton, que es muy pequeña. En otras palabras, cualquier efecto de la gravitación, incluidas las ondas gravitacionales, está suprimido por la extrema debilidad intrínseca de la fuerza gravitacional, algo que

todos experimentamos cuando le ganamos la partida a todo el planeta Tierra mientras sostenemos un lápiz entre nuestros dedos. Así que las ondas gravitacionales requieren eventos de alta energía para generarse significativamente, y además son muy difíciles de detectar directamente. Eso explica los 100 años de espera para detectarlas finalmente en el interferómetro LIGO. Además, ha sido necesario un cataclismo como la colisión de dos agujeros negros de treinta masas solares para generar ondas suficientemente potentes como para que se pudieran detectar aquí, a 1300 millones de años luz de distancia.

Sin embargo, el descubrimiento de las ondas gravitacionales no ha supuesto ninguna sorpresa desde un punto de vista fundamental, pues ya se consideraban prácticamente descubiertas, en un sentido indirecto, desde los años 70, cuando se comprobó que estrellas de neutrones en órbita mutua perdían energía justamente al ritmo adecuado para ser interpretado como una emisión de ondas gravitacionales. Desde un punto de vista de "fundamentos", el descubrimiento más importante sobre la estructura del espacio es sin duda la aceleración de la expansión del universo. Dicho de otra forma, el descubrimiento de la "energía oscura". Desde hace dos décadas sabemos que el 70% de todo el contenido energético del universo está en forma "oscura", es decir, almacenada en el propio vacío.

En sí mismo, el hecho de que el vacío tenga una cierta densidad de energía no es particularmente sorprendente. Después de todo, debería estar "lleno" del campo de Higgs, responsable de las masas de las partículas elementales. Más en general, en cualquier teoría cuántica, el estado de mínima energía tiene necesariamente un mínimo de fluctuaciones, garantizadas por el principio de indeterminación de Heisenberg. La energía de esas fluctuaciones mínimas no se anula, salvo que existan cancelaciones especiales entre diferentes grados de libertad. En la teoría cuántica de partículas elementales, la contribución de fluctuaciones cuánticas a la energía oscura es gigantesca, mucho mayor que la que ha sido efectivamente medida por los astrónomos. Por tanto, sabemos que existe una cancelación misteriosa entre diferentes contribuciones a la energía del vacío, dejando un residuo comparativamente pequeño, pero desconocemos por completo el mecanismo responsable de esta cancelación aproximada.

Cuando Einstein descubrió la relatividad en 1905, demostró que el espacio no es una sustancia ordinaria, porque no tiene sentido preguntarnos si nos movemos o no con respecto al espacio vacío. Sin embargo, sus descubrimientos en la década siguiente contribuyeron a añadir más y más propiedades al espacio: no solo se comporta como un medio elástico, sino que alberga una energía interna perfectamente detectable. Hoy en día, el mayor problema de fundamentos de la física no es tanto la existencia de una energía oscura, como explicar por qué es tan pequeña.

Seguramente, la respuesta a este tipo de cuestiones requerirá una teoría completa de la naturaleza cuántica del espacio-tiempo. En las últimas dos décadas se ha registrado un progreso importante en esta dirección, gracias a la investigación de las propiedades de los agujeros negros. Dado que el espacio tiene propiedades elásticas, nos podemos preguntar si tiene un límite de resistencia a la deformación. En otras palabras, ¿es posible "romper" el espacio mediante la acumulación de suficiente energía localizada? La respuesta es afirmativa: por encima de un cierto límite, el espacio contenido en una cierta región colapsa sobre sí mismo produciendo una singularidad de curvatura infinita. Vista desde fuera, esa región se ve negra, y se denomina "agujero negro".

Gracias al trabajo de Hawking y Bekenstein en los años 70, sabemos que los agujeros negros son la forma más radical de información acumulada en la naturaleza. Sin embargo, es sorprendente que la información oculta en un agujero negro no se organiza en su interior, sino más bien en la superficie del "horizonte de sucesos", el borde exterior del agujero negro. Si esto es así, el interior del agujero negro sería como un holograma: la información estaría localizada en "píxels" en el horizonte, y el interior sería una proyección aproximada. Gerard ´t Hooft y Leonard Susskind propusieron hace dos décadas que todo el espacio podría funcionar de esta forma. En 1997 Juan Maldacena propuso un mecanismo matemático para representar la proyección holográfica en un espacio con energía oscura negativa. Desde entonces, la física teórica ha estado dominada en gran medida por el estudio de este modelo particular.

El funcionamiento del holograma de Maldacena es tan intrincado que una parte importante del trabajo de los físicos teóricos es, literalmente, "desmontarlo" como si de un viejo aparato de radio se tratara, para descubrir los detalles de su estructura interna. ¿Y qué ha emergido como resultado de este esfuerzo? Según parece, el truco fundamental del holograma sería el entrelazamiento cuántico.

El entrelazamiento cuántico es probablemente el aspecto más extraño de la realidad cuántica, tanto que el propio Albert Einstein lo consideró inaceptable. Se dice que dos sistemas cuánticos están totalmente entrelazados cuando su estado físico colectivo está totalmente determinado, pero su estado físico individual está totalmente indeterminado. El sentido común nos diría que para saberlo todo sobre un sistema compuesto es necesario saberlo todo sobre las partes que lo componen. Sin embargo, el entrelazamiento cuántico consiste precisamente en la negación radical de esta idea enraizada en nuestro sentido común. En cualquier caso, por más que se trate de algo poco intuitivo, sabemos que las partículas elementales satisfacen las leyes cuánticas del entrelazamiento.

Los resultados más recientes parecen sugerir que el espacio "está hecho", en un sentido más o menos literal, de entrelazamiento cuántico puro. El ejemplo más espectacular de esto ha sido propuesto recientemente por Maldacena y Susskind, bajo el acrónimo

EPR=ER

Las siglas del miembro izquierdo se refieren a las iniciales de Einstein, Podolsky y Rosen, que describieron las propiedades inesperadas del entrelazamiento cuántico en un famoso artículo en 1935. El mismo año, dos de los autores, en este caso Einstein y Rosen, descubrieron espacios con "agujeros de gusano", atajos espaciotemporales entre dos puntos distantes que conocemos bien por las películas de ciencia ficción. Pues bien, Maldacena y Susskind proponen que dos agujeros negros fuertemente entrelazados estarían unidos por un puente interior, el agujero de gusano de Einstein-Rosen. En este caso, la conexión interior entre los dos horizontes se produciría como resultado de la enorme cantidad de entrelazamiento cuántico entre ambos agujeros negros. A diferencia de los agujeros negros de las películas, de éstos no se puede salir, porque las puertas son agujeros negros, y como es bien sabido, puedes entrar en ellos pero nunca puedes salir.

A día de hoy, la conjetura de Maldacena-Susskind aun espera demostración, incluso a un nivel puramente matemático. Queda para el futuro elucidar si estas ideas

fascinantes nos permitirán avanzar en la respuesta a la pregunta que da título a este capítulo… “¿De qué está hecho el espacio?” La mayoría de nuestras conjeturas matemáticas están basadas en el modelo de Maldacena, que representa la holografía en un mundo con energía oscura negativa. Sin embargo, sabemos que nuestro universo presenta una energía oscura positiva, y en este caso las ideas propuestas no parecen suficientes. Lo más probable es que la solución al problema requiera entender finalmente la enorme cancelación parcial de la energía oscura, algo que, como indicamos anteriormente, se puede considerar como el misterio más urgente de la física fundamental.

Bibliografía:
“La guerra de los agujeros negros. Una controversia científica sobre las leyes últimas de la naturaleza”, Leonard Susskind.
“Geometría y entrelazamiento cuántico”, Juan M. Maldacena (Investigación y Ciencia, Noviembre 2015).
“Black Holes, Information and the String Theory Revolution”, Leonard Susskind and James Lindesay.

José Luis Fernández Barbón
Doctor en Física
Investigador Científico, Instituto de Física Teórica IFT CSIC/UAM

Capítulo 10
¿Por qué se habla de la "baja temperatura del espacio"? ¿Cómo puede tener el vacío una temperatura?

(Por José Oñorbe Bernis)

Efectivamente, en el vacío absoluto no tiene sentido hablar de temperatura. La temperatura de un objeto o un sistema es una medida estadística del movimiento de las partículas que forman ese objeto. Cuanto más rápido se mueven esas partículas, más caliente y cuanto más despacio, más frío. Así que, si no tenemos nada que pueda moverse, obviamente no podemos decir nada de su temperatura.

Lo que sí tiene sentido es pensar en la temperatura que tendría un objeto que pusiéramos en el vacío. Si hiciéramos esto, siempre podremos definir una temperatura midiendo el movimiento de las partículas que forman este objeto. Vamos a ver que la temperatura en el espacio puede ser muy alta o muy baja y que va a depender de donde situemos nuestro objeto. Pero antes de pensar en cómo variará la temperatura de un objeto en el espacio, pensemos primero en un ejemplo más sencillo. ¿Qué pasaría si ponemos un objeto en el vacío?

Imaginemos que en un laboratorio de la tierra tenemos una cámara que nos permite crear un vacío perfecto en su interior [1]. ¿Que pasa si ponemos una esfera de hierro de 1 metro de radio a una temperatura de 23 grados Celsius (°C)? Para poder calcular esto tenemos que pensar en como un objeto puede ganar o perder temperatura y para eso es necesario introducir el concepto de transferencia/transporte de calor. El transporte de calor es la cantidad de energía que se transfiere entre un sistema y su entorno produciendo un cambio de temperatura en dicho sistema. Obviamente este cambio puede ser positivo, aumentando la temperatura, o negativo, disminuyendo la temperatura.

10.1 ¿Cómo puede un objeto cambiar su temperatura? Transferencia de calor.

Existen tres mecanismos posibles de transporte de calor en la naturaleza, que aunque no te suenen por sus nombres ya los conoces muy bien. Estos tres mecanismos son la conducción, la convección y la radiación. La diferencia fundamental entre ellos es como se produce el transporte el calor.

La conducción térmica es el transporte de calor entre regiones de materia que se encuentran a diferente temperatura. El calor es transferido por la interacción directa entre las moléculas y átomos que forman el medio y que se encuentran a distinta temperatura. El ejemplo más claro sucede cuando calentamos una pieza alargada de metal por un extremo y lentamente, por medio de la conducción térmica, el otro extremo del metal también se calentará. Sin embargo en el caso de nuestra esfera de hierro

puesta en el vacío, la conducción térmica no puede jugar ningún papel ¡ya que no se encuentra en contacto directo con nada! Un mecanismo de transferencia de calor descartado.

La transferencia de calor por convección térmica es característica de líquidos y gases e implica movimientos macroscópicos de materia en el sistema. Calentar agua en una olla es un ejemplo muy típico. Una vez la olla se calienta se empieza a transferir calor por conducción al agua en contacto directo con ella. Sin embargo una vez este agua se ha calentado empieza a moverse en la olla, facilitando la transferencia de calor al agua todavía fría. Esta es la idea detrás de los hornos de convección, que mueven el aire del horno facilitando la transferencia de calor por convección. En cualquier caso, volviendo a nuestra esfera en el vacío, está claro que tampoco podemos tener transferencia de calor por convección térmica ya que no se encuentra en contacto con nada.

Nos queda un último mecanismo de transporte de calor y vamos a ver que juega un papel fundamental para entender qué temperatura hace en el espacio: la radiación térmica. La radiación térmica es la transferencia de calor a través de radiación electromagnética. De acuerdo con la teoría cuántica del electromagnetismo, la radiación electromagnética se basa en la existencia de partículas, los fotones, que pueden tener distinta energía y que se propagan a la velocidad de la luz en el vacío. Todo sistema compuesto de materia ordinaria (bariones) es capaz de absorber radiación electromagnética. Por tanto, si apuntamos un láser hacia nuestra cápsula de vacío, la radiación emitida por el láser si puede viajar a través de la cápsula y calentará la esfera de hierro. Un objeto capaz de absorber toda la radiación electromagnética que recibe se conoce como cuerpo negro [2]. Pero en la realidad cada sistema puede ser más o menos eficaz en absorber la radiación (absortividad). Por ejemplo, los hornos microondas están hechos para emitir radiación electromagnética que el agua absorbe muy eficazmente (ondas microondas).

Hemos visto que en el vacío un objeto puede incrementar su temperatura absorbiendo radiación electromagnética pero también disminuirá su temperatura si la emite. ¿Y una esfera de hierro emite radiación electromagnética? Si. Toda la materia ordinaria (bariones) emite radiación electromagnética cuando se encuentra a una temperatura por encima de lo que se conoce como cero absoluto, cero grados Kelvin (ó -273.15 °C). Esta radiación se origina por un conjunto de procesos espontáneos que tienden a incrementar la entropía [3]. La energía por superficie irradiada de esta forma depende principalmente de la temperatura del objeto y del material del que este hecho (emisividad).

Si pudiéramos dejar nuestra cápsula de vacío en un lugar totalmente aislado de cualquier otro emisor, la esfera de hierro irá disminuyendo su temperatura conforme emite radiación electromagnética. Por ejemplo, la temperatura de nuestra esfera de hierro bajará de 23°C a 22°C, en unas tres horas y quince minutos [4]. Un ser humano de tamaño medio (1.70 metros y 70 Kg de peso) a 36°C tardaría unos 5 minutos en enfriarse un grado por el mismo proceso físico y una media hora en llegar a los 30°C, límite de la hipotermia grave y la pérdida de conciencia [5]. Sin embargo, en la realidad mucho antes de que pudiéramos sentir frío, la exposición del cuerpo humano a un medio sin presión como es el vacío sería un problema mucho más grave.

10.2 La temperatura en el espacio

Ahora ya sabemos que el vacío puede ser un sitio muy frío pero solo si estamos realmente solos, es decir, si no recibimos radiación electromagnética. Pero esto no tiene porqué ser así en el espacio. Las estrellas ó los quásares son dos ejemplos de fuentes de radiación electromagnética en el Universo. De hecho el espacio puede ser un sitio muy caliente si estamos cerca de una estrella. Por ejemplo, si ponemos nuestra esfera de hierro en el espacio cerca la estación espacial internacional de tal forma que reciba directamente la radiación del Sol, solo hacen falta unos ochenta minutos para que la temperatura de nuestra esfera se incremente en un grado. La esfera alcanza los 60.35°C, que es su temperatura de equilibrio, la temperatura a la cual el objeto emite la misma radiación que absorbe, en unos veinte días [6]. Por tanto, las diferencias de temperatura entre un objeto que recibe radiación solar y otro que no lo hace son muy grandes. Un ejemplo muy significativo es la Luna, la diferencia de temperatura en la superficie de la Luna entre el día y la noche llega a alcanzar los doscientos grados centígrados (de los -153°C durante la noche a 107°C durante el día). Los astronautas también sufren estas diferencias de temperatura, pero los trajes espaciales además de estar fabricados para reflejar casi el 90% de la luz que les llega, y por tanto tener una absortividad muy baja, llevan incorporados sistemas de refrigeración.

Conforme nos alejemos de la fuente de radiación, la temperatura de un objeto se irá reduciendo. Sin embargo aún intentando alejarse de toda fuente posible, nunca veríamos a un objeto alcanzar una temperatura de equilibrio de cero grados Kelvin sino 2.7 Kelvin (-270.45 °C). Esta es la temperatura de la radiación de fondo de microondas producto del Big Bang [7] y que baña por completo el Universo. Así que esta sería la temperatura que tendría nuestra esfera de hierro si consiguiéramos aislarla del resto de fuentes del Universo. Por cierto, ¿cuánto tardaría en alcanzar la esfera esta temperatura? ¡unos 80.000 años! [8] Aunque esto parezca mucho tiempo, es bueno recordar que el sistema solar tiene unos 4.600.000.000 años (y el Universo tiene aproximadamente unos 13.799.000.000 años). Con esto en mente, está claro que el tiempo no es un problema para que la mayoría de los objetos que forman el Universo alcancen una temperatura de equilibrio. Quizás por eso no es extraño entonces que mucha gente piense en el espacio como un lugar frío. Pero esto es solo cierto siempre y cuando no tengamos ninguna fuente de calor cerca, y nosotros tenemos el Sol.

Notas:
1. Estas cámaras de vacío existen en la Tierra aunque suelen construirse con materiales que las aíslan perfectamente del exterior. La cámara de vacío más grande está en Estados Unidos (Sandusky, Ohio) y es un cilindro de 37 metros de diámetro y 37 metros de altura. Fue construida por la NASA en 1967 y se ha usado para probar distintas partes de cohetes, de módulos de aterrizaje de Marte y de estaciones espaciales.
2. Ver capítulo 62.
3. Ver capítulo 58.
4. Para hacer este cálculo hemos asumido una emisividad constante para el hierro de 0.2 (hierro pulido), una capacidad calorífica de 412 J $Kg^{-1}K^{-1}$ y una densidad de 7850 kg m^{-3}.
5. Para hacer este cálculo hemos asumido una emisividad constante para el tejido humano de 0.99 y una capacidad calorífica de 3500 J $Kg^{-1}K^{-1}$. Hemos calculado la superficie total del cuerpo humano usando la relación empírica hallada por Mosteller (ver bibliografía) entre esta cantidad y la altura y el peso de la persona.
6. Hemos considerado que la superficie del Sol está a 5800 Kelvin, que el Sol tiene un radio de $6.995x10^{8}$ metros y que nuestra esfera esta a $1.50x10^{11}$ metros del Sol, que es la misma que del Sol a la tierra. Además hemos asumido que la esfera de hierro tiene una absortividad de 1.0, es decir que es capaz de absorber toda la radiación electromagnética como un cuerpo negro perfecto. Esto no es del todo cierto, ya que parte de esa radiación será reflejada.
7. Ver capítulo 59.
8. La temperatura del fondo cósmico de microondas, 2.725 Kelvin, nos da una intensidad de radiación por superficie de $3.01x10^{-6}$ W/m^2. De nuevo, hemos asumido que la esfera de hierro es un cuerpo negro perfecto y que tiene una absortividad de 1.0.

Bibliografía:
- “Física: Mécanica, radiación y calor”, R.P. Feynman, R.B. Leighton, M. Sands (1987). Adison Wesley Iberoamericana S.A.
- “Statistical Physics”, L.D. Landau, E.M. Lifshitz (1996). Oxford: Butterworth–Heinemann.
- “Simplified calculation of body-surface area”, RD. Mosteller (1987). N Engl J Med 317:1098.
- http://www.nasa.gov.
- http://www.esa.int
- http://www.universetoday.com/
- Todas las emisividades para distintos materiales fueron tomadas de https://en.wikipedia.org/wiki/Heat_capacity
- Todas las capacidades calorífica para distintos materiales fueron tomadas de http://www.optotherm.com/emiss-table.htm

José Oñorbe Bernis
Doctor en Física
Max Planck Institut für Astronomie

Capítulo 11
¿Qué es el polvo cósmico y de dónde viene?
(Por Ricardo Dorda Laforet)

La cultura popular suele considerar que el espacio que se extiende entre las estrellas es un gran vacío, pero en realidad en este espacio hay grandes cantidades de gas y polvo, que reciben el nombre de medio interestelar. Ahora bien, la densidad del medio interestelar es muy baja en comparación con lo que estamos habituados (nuestra propia atmósfera), tan baja que es incluso menor que lo que podemos llegar a obtener artificialmente en los laboratorios. En ese sentido, efectivamente podemos considerar que los abismos interestelares están cuasi vacíos, pero estas regiones son tan inimaginablemente grandes que, pese a su baja densidad de materia, cualquiera de ellas puede albergar nubes de gas y polvo con masas de cientos o miles de veces la del Sol.

11.1 Analizar lo invisible

Como ya se ha mencionado, el medio interestelar está formado por gas y polvo, pero no fue fácil percatarse de ello. Al contrario que las estrellas, esta materia no emite luz visible en condiciones normales, por lo que resultaba invisible para los astrónomos. Sin embargo, que fuera invisible no significa que no fuera detectable. La radiación electromagnética (luz) que pasa a través del medio interestelar se ve afectada por el gas y el polvo en él, que dispersan y absorben la luz de la misma forma que en la Tierra una nube puede ocultarnos la luz del Sol. Dada la baja densidad de esta materia interestelar, podría parecer difícil que bloquee la luz de una estrella, pero hay que tener en cuenta su enorme tamaño. Así, son tan profundas que pueden llegar a ocultarnos completamente la luz de las estrellas que están al otro lado.

Además, el medio interestelar no afecta a todas las longitudes de onda de la luz [1] con la misma intensidad. Así, el medio interestelar afecta sobre todo al rango de la luz visible, pero más a las longitudes de onda más cortas (el ultravioleta cercano y color azul) que a las más largas (el color rojo o el infrarrojo). En consecuencia, cuando el medio interestelar bloquea parte del brillo de las estrellas, además provoca que las veamos más rojas de lo que en realidad son. Los astrónomos de la segunda mitad del Siglo XIX ya se habían percatado que algunas estrellas eran más débiles y más rojas de lo que les correspondía, pero no fue hasta 1930 que Trumpler dedujo que esos efectos se debían al gas y polvo entre las estrellas y nosotros. Los científicos empezaron entonces a estudiar la composición química de esta materia de forma indirecta, a partir de su efecto en la luz de las estrellas que la atraviesan. Con el desarrollo de la radioastronomía a lo largo del Siglo XX, los astrónomos también pudieron estudiar la composición del medio interestelar mediante su observación directa, ya que el gas y el polvo interestelares emiten radiación electromagnética de baja frecuencia, en el rango desde las microondas a las ondas de radio.

Estudiando la forma en que el medio interestelar absorbe la luz de las estrellas y la forma que emite radiación electromagnética los científicos han podido desvelar su composición y propiedades, aunque aún hoy en día se investiga intensamente en este área. Por un lado, el polvo consiste en pequeñas motas sólidas cuyo tamaño típico va desde unas pocas micras (la millonésima parte de un metro) a la décima parte de eso, y cuya composición es muy variada. La componente principal del polvo son silicatos, es decir lo mismo que la arena de los desiertos, aunque los granos del polvo interestelar son unas mil veces más pequeños que los granos de arena terrestres. Además de silicatos, también hay motas de otros agregados formados por otros metales o sus óxidos (como óxido de aluminio), por no metales como carbono (grafitos) o incluso hielos de agua o de monóxido de carbono. Por otro lado, el gas se compone mayormente de hidrógeno atómico y molecular, con algo de helio y trazas de muchos otros elementos y sustancias, incluso moléculas orgánicas de cierta complejidad como fullerenos, “pelotas” formadas por 60 átomos de carbono.

11.2 Astroquímica

Después de leer los párrafos previos, es muy probable que el lector se pregunte de dónde viene toda esa materia que forma el medio interestelar y su variedad química. La cosmología ha determinado que en el Universo primigenio, antes de que las primeras estrellas nacieran, la materia convencional era un 75% Hidrógeno, un 25% Helio y una pequeña fracción constituida por isótopos de estos elementos y Litio. Cuando nacieron las primeras estrellas en sus núcleos comenzó la fusión del Hidrógeno y el Helio en elementos progresivamente más pesados [2]. Cuando esta primera generación estelar murió, liberaron al espacio todo tipo de elementos químicos que se mezclaron con el resto del gas allí presente. Las siguientes generaciones de estrellas nacieron de este gas y con sus muertes enriquecieron el medio interestelar con elementos aún más pesados, en un proceso que continúa aún hoy en día y que proporciona los ingredientes para la variedad de sustancias químicas y el polvo presente en el gas interestelar.

Ahora bien, como se ha mencionado antes, la densidad del medio interestelar es bajísima. Tan baja que puede haber solo un átomo por metro cúbico, por lo que la probabilidad de que dos partículas choquen es extremadamente pequeña. En esas condiciones, ¿cómo es posible que ingentes cantidades de átomos terminen aglomerándose para dar lugar a una mota de polvo? Más aún, ¿cómo pueden formarse ciertas moléculas químicas, especialmente las más complejas, y sobrevivir a las agresivas condiciones del medio interestelar?

La respuesta a la primera pregunta es relativamente sencilla: el polvo no nace en las nubes interestelares, sino en la atmósfera de ciertas estrellas moribundas, de dónde luego es liberado al medio interestelar. Veámoslo con más detalle. La temperatura superficial de la gran mayoría de estrellas es demasiado alta como para que los elementos pesados puedan formar moléculas ni mucho menos motas de polvo. Sin embargo, cuando la gran mayoría de estrellas (todas aquellas con masas entre un décimo y treinta veces la del Sol [3]) terminan la fase principal de su vida, evolucionan expandiéndose enormemente. Se vuelven decenas e incluso cientos de veces más grandes de lo que eran originalmente, al tiempo que su superficie se enfría considerablemente. Se convierten así en las estrellas que llamamos gigantes o supergigantes rojas. Estas estrellas poseen atmósferas frías que, debido al tamaño de la estrella, están solo débilmente ligadas a su estrella y es habitual que parte de ellas se

vaya desprendiendo lentamente hacia el medio interestelar, especialmente en las últimas fases de la vida de la estrella.

Es en estas capas superiores de la estrella y en las nubes que se han desprendido de ella donde se dan las condiciones ideales para la formación de motas de polvo. La temperatura resulta suficientemente baja y la densidad suficientemente alta durante tiempos suficientemente largos (miles de años) como para que los átomos se vayan agregando en motas cada vez mayores. Al mismo tiempo, la estrella está suficientemente fría como para que su radiación no sea tan energética que destruya los incipientes granos de polvo. Con el tiempo, estas nubes de gas y polvo se alejan progresivamente de la estrella en la que han nacido, dispersándose en el espacio y dando lugar al medio interestelar.

Por otro lado, para responder a la segunda pregunta hay que hablar de la simbiosis que existe entre las motas de polvo cósmico y las moléculas que se forman en las nubes de gas. Muchas reacciones químicas requieren condiciones particulares para ser eficientes. De nada sirve la presencia de los ingredientes de una molécula, si no se dan las condiciones adecuadas. Tal vez el ejemplo más fácil para ilustrar esto sea la molécula de Hidrógeno, formado por dos átomos de este elemento. La probabilidad de que se forme esta molécula a partir del impacto de dos átomos libres de Hidrógeno resulta absurdamente pequeña incluso en condiciones de densidad favorables. De hecho, por experimentos en laboratorio sabemos que estas moléculas se forman a partir de impactos triples: en el choque dos de las moléculas transmiten suficiente energía cinética a una tercera como para quedar ligadas entre ellas. Este impacto triple puede resultar relativamente frecuente en altas densidades, pero definitivamente no puede explicar la presencia de Hidrógeno molecular en las extremadamente difusas nubes interestelares.

La respuesta a este misterio se halla en el polvo interestelar que acompaña al gas. Se sabe que los átomos de Hidrógeno tienden a adherirse fácilmente a la superficie de los granos de polvo. Entonces, debido a la adherencia de múltiples átomos de Hidrógeno, es fácil que se forme hidrógeno molecular en la superficie del grano de polvo. Es decir, el polvo actúa como un catalizador químico en esta y en muchas otras reacciones, pero su papel aún va más allá. Hay que tener en cuenta que las moléculas son relativamente frágiles ante el impacto de rayos cósmicos o de radiación de alta energía, cosas que abundan en el espacio interestelar y que destruyen progresivamente las moléculas. Los granos de polvo, más resistentes, actúan como escudo para las moléculas. Así, en las nubes más compactas y frías, el polvo de la parte más exterior protege a las moléculas del interior de la nube, disminuyendo la velocidad a la que son destruidas por el entorno. En definitiva, el polvo cósmico es indispensable para explicar la variedad molecular observada en el medio interestelar.

11.3 Semillas de vida

Con el tiempo, una nube de gas y polvo del medio interestelar empezará a colapsar sobre si misma. Esto puede ocurrir inducido por su propia gravedad, pero también por factores externos, como verse afectadas por la radiación o el viento estelar de estrellas luminosas, la presencia de intensos campos magnéticos producidos en objetos como púlsares o agujeros negros, o el encontronazo con el repentino destello de una supernova o su posterior onda expansiva. El colapso de una nube provoca que se

vaya fragmentado en grumos, cada uno de ellos con sus propios puntos de mayor densidad, donde empiezan a nacer nuevas estrellas.

El gas y el polvo cercano a una estrella en formación cae lentamente hacia ella, formando en el proceso un disco alrededor suyo. Es en este disco dónde los grumos de gas empiezan a atraer hacia sí la materia cercana a ellos en su órbita alrededor de la estrella, lo que da lugar a los planetas gaseosos gigantes. Pero más aún, es el polvo el que agregándose en partículas cada vez mayores forma pequeñas rocas, que a su vez se aglomeran para formar planetesimales: cuerpos sólidos cuya unión dará lugar a los planetas rocosos. Así pues, sin polvo cósmico no sería posible la formación de planetas como el nuestro.

Con el tiempo la acción combinada de la atracción gravitatoria de los planetas y la radiación de la estrella terminan limpiando de gas el disco alrededor de ésta, quedando tan solo planetas con sus lunas y algunos escombros, como nuestro cinturón de asteroides. En la parte más externa del sistema planetario (como es el caso de nuestro cinturón de Kuiper) pueden acumularse muchos de estos escombros, producto de la acreción de polvo y hielos de sustancias volátiles, y que contienen muchas sustancias químicas que estaban presentes en la nube original. Son estos cuerpos los que, cuando por cualquier razón caen hacia el interior del sistema planetario, se convierten en cometas y sus hielos volátiles y las sustancias que albergan éstos pueden llegar a la superficie de los planetas ya formados. Así, los científicos sospechan que algunos de las moléculas orgánicas necesarias para el origen de la vida llegaron en los cometas. Y es que es difícil que en la química de la Tierra primigenia se formaran ciertas sustancias orgánicas, pero se sabe que estas moléculas existen en el medio interestelar. Por tanto, es posible, aunque aún no existe consenso, que esos ladrillos de la primera vida se hubieran formado mucho antes que la Tierra, en atmósferas de estrellas moribundas o en frías nubes interestelares, pero en cualquier caso gracias al papel del polvo cósmico en la química interestelar.

Notas:
1. Ver capítulo 66.
2. Ver capítulo 15
3. Ver capítulo 17

Ricardo Dorda Laforet
Doctor en Astrofísica
Investigador post-doctoral en la Universidad de Alicante

Capítulo 12
¿Qué son los pulsares? ¿Y los quásares?
(Por Paola Marziani)

Suena extraño asociar pulsares y quásares en el mismo capítulo, ya que son objetos muy diferentes. Los quásares fueron descubiertos más de diez antes del momento en el que Isaac Asimov publicaba su recopilación sobre las preguntas más básicas sobre la ciencia. El descubrimiento de los quásares fue reportado en 1963 por Marteen Schmidt en la revista Nature. Por otro lado, los pulsares fueron descubiertos poco tiempo después en 1967 por Jocelyn Bell Burnell y Antony Hewish. Tanto los quásares como los pulsares habían entusiasmado la mente, acrecentando la fantasía, y las esperanzas de mucha gente fuera del mundo de los astrónomos profesionales. Hacia finales de los años 60's los quásares seguían siendo demasiado misteriosos para ofrecer algunas respuestas básicas acerca de sus propiedades ya que las hipótesis sobre su origen se contraponían entre sí. El mismo Isaac Asimov se preguntaba:

"Y también están los quásares, que según algunos son un millón de estrellas ordinarias, o más, colapsadas todas en una ¿Qué decir de la temperatura de su núcleo interior? Hasta ahora nadie lo sabe."

¿Pero qué son los pulsares y los quásares? Antes de discutir en detalle los aspectos físicos y presentar algunas analogías recordemos sus propiedades más básicas, dedicando unas palabras sobre el contexto histórico de su descubrimiento. Los pulsares son objetos que emiten impulsos de radio a una frecuencia muy estable. La frecuencia es diferente para cada pulsar, varía de 1 a 30 ciclos por segundo, de 1000 a 0.5 ciclos por segundo para todos los pulsares conocidos. Se detectaron por primera vez con observaciones de radio, como fuentes débiles que fueron pasadas por alto en los primeros mapas. Al principio, la regularidad y la frecuencia de los pulsos de radio (el término pulsar es una contracción de estrella pulsante, pulsating star, en inglés) sorprendió a los observadores. En ese momento, la radioastronomía era una ciencia relativamente joven, sin embargo la gente empezaba a disfrutar la explotación sin precedentes de las comunicaciones a través de ondas de radio, desde los aparatos de radio portátiles de transistores ("pocket radios" en inglés, "radioline" en italiano) hasta los aparatos televisivos. Estos dispositivos eran un signo del avance de nuestra civilización. El poder intrínseco, la frecuencia y la regularidad de los pulsares presentaban un reto para los procesos físicos que ya eran bien comprendidos en el momento. ¿Y por qué no pensar que esas señales tan regulares eran debidas a una comunicación de una civilización extraterrestre? Es comprensible que esta idea no parecía tan ingenua y de fantasía como parece ahora. La década de los 60's fue un momento en que nuestra visión del mundo y del Universo estaba cambiando radicalmente: los primeros seres humanos en el espacio, la cara oculta de la luna observada por la primera vez, las galaxias descubiertas a distancias cada vez mayores con más potentes telescopios. El universo estaba apareciendo también por nuevas

ventanas en el espectro electromagnético (radio y rayos X) que revelaban la presencia de fenómenos sumamente energéticos y violentos. El concepto de “cielos inmutables” de la antigüedad (considerado como un dogma hasta el siglo XVII) estaba siendo anulado para siempre. El universo parecía más familiar y accesible; sin embargo, las distancias se hacían más grandes cada día. El descubrimiento de los quásares añadió una importante contribución a esta visión ampliada del Universo. Pronto se entendió que los quásares estaban a distancias muy grandes deducidas de sus desplazamientos al rojo según la ley de Hubble, que implicaba la formidable distancia de 1.5 mil millones de años luz en el caso de 3C273, el primer quasar (para un valor de la constante de Hubble de 100 kilómetros por segundo por megaparsec). Su descubrimiento fue pronto seguido del descubrimiento de quásares de corrimiento al rojo tan grande como 2, que implica una distancia de 10 mil millones de años luz, con luminosidades que parecían increíbles.

¿Cuál es el mecanismo físico que da lugar a un pulsar? La fuente debe ser pequeña debido a la frecuencia del pulso (más pequeño que un segundo luz, y mucho más pequeño que las estrellas ordinarias), debe ser muy estable y tiene que haber algún mecanismo capaz de producir intensas emisiones de radio. Tal mecanismo no puede ser de naturaleza térmica, es decir, no puede ser debido a la energía térmica de un cuerpo caliente, como en una estrella. Un mecanismo físico adecuado es la radiación de sincrotrón emitida a través de la aceleración de los electrones y otras partículas cargadas a velocidades relativistas en un campo magnético fuerte.

Un pulsar es una estrella compacta en rotación (el diámetro debe ser del orden de 10 km) con un campo magnético fuerte, y con el eje del campo desalineado respecto al eje de rotación. El campo magnético en un cuerpo en rotación produce un campo eléctrico que acelera las partículas cargadas cercanas a la superficie de la estrella a velocidades muy altas, produciendo un haz de radiación electromagnética muy intenso a lo largo del eje magnético (en realidad, el mecanismo que produce esta radiación no es bien conocido). Por otro lado, el eje magnético no está alineado con el eje de rotación, esto ocurre también con otros cuerpos celestes, como por ejemplo nuestro planeta Tierra.

Los pulsos de radiación serán visibles para un observador terrestre solamente cuando el eje está alineado con la línea de visión (como por las lámparas de los faros en rotación). Hay un límite inferior de la densidad de una estrella para estar unida gravitacionalmente, ya que la aceleración centrífuga en su ecuador no podrá ser superior a la aceleración de la gravedad. Para la velocidad angular de los pulsares, la densidad debe ser extremadamente alta, tan alta que debe ser comparable a la densidad de la materia nuclear. Cuando una estrella masiva de más de 8 veces la masa de nuestro Sol sufre las últimas fases de su evolución, detona como una supernova. Las capas externas son expulsadas al espacio, pero el núcleo de un par de veces la masa del sol se contrae por su gravedad. La presión gravitacional es tan fuerte que supera la repulsión electrostática que mantienen estables a los iones y núcleos; de este modo, los electrones y protones se funden en neutrones. El colapso es entonces detenido por la presión asociada con los neutrones. Si la masa final es de más de 1.4 masas solares (este valor es llamado el límite de Chandrasekhar), el colapso puede progresar para formar un objeto totalmente colapsado, una estrella de neutrones; y si es de más de 3 masas solares, entonces se formará un agujero negro. Durante el colapso de la estrella progenitora, el diámetro de la estrella se reduce de forma drástica, pero se conserva el momento angular (¡no hay fuerzas externas que ejercen una torsión en el sistema!), de modo que la velocidad de rotación se incrementa dramáticamente. Es el mismo caso de

una bailarina que sigue girando más lentamente con los brazos extendidos, y más rápidamente con los brazos mantenidos cerca de su cuerpo.

El remanente del colapso es una estrella muy compacta, extremadamente densa (tan densa como los núcleos atómicos), estable y de rotación rápida, cuyos campos magnéticos actúan como una especie de imán giratorio: un pulsar. La rotación del pulsar es tan estable que su regularidad compite con la estabilidad de los relojes atómicos, por ejemplo con algunos pulsares con períodos en el intervalo de 1 a 10 milisegundos. Sin embargo, la energía del sistema radia lentamente, y si el ambiente interestelar de la estrella de neutrones es suficientemente rico, la estrella puede ser frenada por la fricción ejercida sobre su campo magnético. Con el tiempo la estrella puede girar más lentamente hasta que se alcanza un estado de reposo.

Un pulsar "gastado" puede regresar a la vida, si es parte de un sistema binario. Si su compañera estelar llega al final de su vida en la secuencia principal, comenzará a expandirse. Parte del material de esta compañera puede caer sobre el pulsar, poniéndolo de nuevo en rotación. El proceso es, obviamente, favorecido en ambientes estelares densos, como los cúmulos globulares. En el núcleo de un cúmulo globular, la distancia típica entre las estrellas es de 10 a 100 veces menor que la distancia típica en la vecindad solar. Las estrellas están tan cerca que se pueden formar nuevos sistemas binarios.

Cuando una estrella de neutrones se une con una nueva estrella compañera, su fuerza de gravedad es tan intensa que puede extraer el gas de la compañera. Este gas es finalmente acretado por la estrella de neutrones, rejuveneciendo el pulsar, ya que el momento angular del gas se transfiere desde la compañera a la estrella de neutrones.

Los pulsares son fácilmente identificables en la Galaxia, y cada pulsar tiene su frecuencia individual. Es fascinante que un mapa con la posición del Sol en relación con 14 pulsares conocidos haya sido incluido en la nave espacial Pioneer para permitir que posibles seres extraterrestres inteligentes puedan rastrear la posición del sistema solar en el espacio y el tiempo. La frecuencia de los pulsos emitidos es tan estable, que una gran variedad de pulsares está siendo actualmente utilizado como detector de ondas gravitacionales debido a que se espera que estas ondas induzcan un cambio en la frecuencia de los pulsos.

Las estrellas de neutrones son objetos pequeños y masivos cuyo campo gravitacional es extremadamente fuerte: lo que pesa 1 g de materia en la Tierra pesaría más de cien millones de toneladas en la superficie de una estrella de neutrones con masa al límite de Chandrasekhar, 1.4 $M_{\odot}$. Es decir, ¡una vez y media la masa del sol está comprimida en tan solo 10 kilómetros, el radio típico de una estrella de neutrones! Por lo tanto, los pulsares tienen terreno para probar la teoría de la relatividad general de la gravitación.

El púlsar binario PSR1913+16 es un sistema estelar en el que ambos componentes son pulsares. Este sistema ha proporcionado una de las confirmaciones más rigurosas de la relatividad general, ya que sus órbitas se van desgastando según lo predicho por la teoría, a través de la emisión de radiación gravitacional.

Los quásares son fuentes muy diferentes de los pulsares. La interpretación aceptada actualmente por la mayoría de los científicos implica acreción, que es

literalmente caída de materia sobre un objeto compacto, muy probablemente un agujero negro súper masivo. La acreción sobre un objeto compacto es un proceso tremendamente eficiente, en donde la “eficiencia” se refiere a la producción de la radiación. Al igual que los niños que deslizan por una pendiente para adquirir velocidad debido a la pérdida de energía potencial que se convierte en energía cinética, el gas que cae en un agujero negro también adquiere energía cinética. Esta energía es disipada por la fricción viscosa de los gases sobre sí mismos, al igual que las ruedas de un coche cuando el conductor frena. El gas se calienta y comienza a brillar intensamente. El potencial gravitacional es como un pozo profundo para el caso de un objeto compacto como un agujero negro. El gas que está cayendo sigue una “pendiente” muy empinada hasta el pequeño tamaño del objeto compacto. Por esto, la cantidad de energía liberada es una fracción significativa de la masa en reposo de gas, casi el 7% para un agujero negro no giratorio. De acuerdo con la ecuación fundamental de Einstein que da la equivalencia entre masa y energía, la energía es igual a la masa multiplicada por el cuadrado de la velocidad de la luz. No es de extrañar que los quásares puedan ser extremadamente luminosos llegando a tener luminosidades de 1038 kilowatt, es decir, de 100 a 1000 veces más luminosos que una galaxia gigante. Los agujeros negros tienen que ser muy masivos, de lo contrario la fuerza ejercida por la radiación alejaría el gas que proporciona la luminosidad del quasar. Las masas de la mayoría de los quásares son mayores a un millón de veces la masa solar, o lo que es lo mismo, la masa de una galaxia enana. Los agujeros negros más masivos que se conocen pueden alcanzar a los diez mil millones de masas solares, es decir, ¡la masa estelar de una galaxia gigante!

Toda la masa está concentrada en un tamaño mínimo, contenido en el radio gravitacional del agujero negro: para una masa de 100 millones de masas solares, el tamaño sería de 150 millones de kilómetros, esto es, la distancia entre la Tierra y el Sol. Los agujeros negros súper masivos habitan en los núcleos de las galaxias. Inclusive, hoy en día se cree que estos agujeros negros están presentes en galaxias más masivas como espirales y elípticas gigantes. Con mucha frecuencia en las galaxias cercanas, algunos se encuentran en reposo y otros están acretando, creando núcleos luminosos en sus galaxias anfitrionas.

Desde la década de los 70’s, los quásares han perdido un poco de su “brillo exótico”. Los astrónomos se dieron cuenta de que los núcleos de las galaxias cercanas también poseían la acreción de agujeros negros, desde entonces se ha reconocido a una clase más general de "núcleos activos de galaxias” (en inglés, Active Galactic Nuclei - AGN). Hoy en día la palabra quasar está fuera de moda, el término AGN o agujeros negros súper masivos se usan como sinónimo desde algún tiempo. Sin embargo, la definición de los quásares como AGN luminosos no tiene una definición muy clara debido a que abarcan una enorme gama de luminosidades. Astronómicamente, los quásares son objetos débiles (el más brillante tiene magnitud aparente de 13), relativamente raros (alrededor de 100 por grado cuadrado hasta magnitud 22, y ~20 hasta magnitud 20).

Son fáciles de distinguir por su color azulado y su espectro de abundantes líneas de emisión (los espectros son similares pero no idénticos para todos los quásares). Una propiedad que desconcertó a los descubridores de los quásares era que no podían ver la galaxia madre. Pero pronto se entendió que el tamaño angular de la galaxia anfitriona era demasiado pequeño para los objetos distantes y demasiado débiles para ser detectada.

Los quásares emiten una enorme cantidad de radiación en todas las frecuencias del espectro electromagnético, desde ondas de radio hasta los rayos X duros. Su espectro es consistente con una emisión “no térmica” como en el caso de los pulsares, aunque la formación del continuo de los quásares está asociada a diferentes regiones. Se cree que la mayor parte de las emisiones ópticas y UV provienen del disco de acreción caliente el cual emite un espectro de naturaleza térmica. Sin embargo, la emisión en ondas de radio no puede ser térmica y es debida a radiación de sincrotrón.

Lo que hace a los quásares especialmente interesantes es que son trazadores de un pasado lejano en la historia del Universo. Los quásares más luminosos se observan a un corrimiento al rojo $z > 1$ cuando la edad del Universo era de alrededor de 1/3 de la época actual, por lo que la población de los quásares (~ 300.000 se conocen hasta hoy) evoluciona a través de la edad cósmica. Hoy en día, los núcleos de las galaxias están poblados por agujeros negros muy masivos que reciben sin embargo muy poco material de acreción. Probablemente, las galaxias masivas fueron quásares en su pasado distante.

Actualmente, las galaxias en el Universo local están descansando, ya que la gloria de cuando fueron quásares brillantes pasó hace mucho tiempo.

Los agujeros negros súper masivos en los núcleos de las galaxias necesitan mucha materia en acreción. Para ser muy luminosos, deben consumir más de 10 veces la masa del Sol por año. Algunos quásares locales pueden ser rejuvenecidos, como los pulsares que reciben material de acreción, pero para los quásares este fenómeno es relativamente raro. Algunos quásares pueden estar hechos de un par de agujeros negros con una pequeña separación orbital, formando un sistema binario como los de estrellas pero de agujeros negros. Sin embargo, se espera que la transferencia de materia entre ellos sea diferente a las binarias estelares, donde el compañero de la estrella de neutrones está todavía en evolución y produce un viento. Un mecanismo eficaz en el suministro de combustible para alimentar literalmente al monstruo en los núcleos de las galaxias, es la interacción gravitacional entre galaxias. Las fuerzas de marea pueden transportar una enorme cantidad de gas hacia el centro de las galaxias, y si hay un agujero negro allí, éste crece y su luminosidad crece con él. Los encuentros entre galaxias pueden llevar a una fusión literal entre ellas, esto es, que dos galaxias se fusionen y pierdan su apariencia individual, dejando como resultado a una galaxia remanente semejante a una galaxia elíptica. Estos fenómenos eran mucho más frecuentes en el pasado, en la llamada época cósmica. Por lo que es una sorpresa que el primer quásar que se descubrió tuviera un corrimiento al rojo de ~ 0,15, lo que implica una distancia de dos mil millones de años luz con los parámetros cosmológicos actuales, y que los quásares más brillantes se encuentran a un corrimiento al rojo de ~ 1 – 2, cuando la edad del Universo era entre 3 y 6 mil millones de años.

Los pulsares y los quásares provocaron una revolución cultural que marcó el comienzo de la astrofísica en el dominio de los fenómenos no térmicos, los cuales son procesos físicos fundamentalmente diferentes de los que alimentan a las estrellas. Siguen siendo temas de frontera hasta la actualidad. Los quásares y los pulsares son activados por objetos compactos (estrellas de neutrones y agujeros negros) que todavía no están bien comprendidos. Cuanto más nos acercamos a las estrellas de neutrones y a los agujeros negros, menos conocemos de su física. Siguen siendo problemas abiertos los relacionados a la aceleración del chorro del púlsar que emite la intensa luz colimada, a la naturaleza del horizonte de eventos de los agujeros negros, la interacción agujero negro con el medio circundante y su campo magnético. Además, las estrellas de

neutrones y los agujeros negros son objetos donde la fuerza de gravedad es extremadamente fuerte. Son laboratorios en los que la principal teoría de la gravitación, la relatividad general se puede probar con éxito.

Se agradece la ayuda de las doctoras **Alenka Negrete** y **Alba Grieco** por una revisión precisa de este capítulo.

Bibliografia:
M. D'Onofrio, P. Marziani, J. W. Sulentic, Fifty years of quasars, Berlin:Springer (2012)
D. Dultzin, Quásares/ Quasar: En Los Confines Del Universo

Paola Marziani
Doctora en Astrofísica
Investigadora / INAF - Osservatorio Astronomico di Padova

Capítulo 13
¿Qué es un agujero negro? ¿Tiene limitaciones de tamaño?

(Por Luis Julián Goicoechea Santamaría)

La idea de agujero negro está directamente relacionada con la ley de la gravitación universal que propuso Isaac Newton en el siglo XVII, en la cual se establecía que un objeto con cierta masa ejerce una fuerza atractiva sobre otros objetos, cuya intensidad es inversamente proporcional al cuadrado de la distancia. Así, la fuerza gravitatoria aumenta a medida que la distancia al atrayente (con una masa dada) disminuye. Aunque la gravedad de un planeta o una estrella cerca de su superficie es importante, los objetos pueden contrarrestar dicha atracción y escapar si disponen de suficiente energía. Por ejemplo, la velocidad de escape desde la superficie terrestre es de unos 11 kilómetros por segundo, mientras que se necesita una velocidad de 617 kilómetros por segundo para escapar de una estrella como el Sol. Nosotros sabemos que la radiación (y las partículas más energéticas) no pueden ser capturadas por el campo gravitatorio solar, ya que la luz visible que proviene del Sol (velocidad de 300000 kilómetros por segundo) ilumina el cielo diurno.

Sin embargo, si un objeto colapsa hasta introducirse dentro de un radio extremadamente pequeño, en el cual la velocidad de escape es igual a la velocidad de la luz, se forma un agujero negro. La materia y la radiación quedan atrapadas en su campo gravitatorio, y no son capaces de propagarse hasta otras regiones vecinas. Dicho radio crítico suele definir el tamaño del agujero negro, aunque más que un tamaño real es una frontera, en cuyo interior se ha formado la mencionada estructura compacta y oscura. El radio crítico (o tamaño radial) es directamente proporcional a la masa, y es de unos 3 kilómetros para la masa del Sol, y de solo 9 milímetros para la masa de la Tierra. En otras palabras, para transformar la Tierra en un agujero negro, necesitaríamos comprimir toda su masa en un volumen menor que el de un ojo humano.

Desde un punto de vista teórico, se especula con la posible existencia de agujeros negros primordiales, formados mediante fluctuaciones locales en la densidad de materia durante la expansión inicial del Universo. Se trataría de objetos poco masivos y exóticos, que pueden tener escapes a través de sus fronteras (radiación de Hawking), de forma que no se comportarían como cárceles totalmente eficientes para la materia y la radiación. Estos escapes estarían relacionados con efectos cuánticos, y generarían una disminución paulatina de su masa hasta la evaporación de los mismos. En los últimos años, Stephen Hawking también ha postulado que la materia y la radiación no colapsan dentro de la esfera con el radio crítico del agujero negro, ya que quedarían atrapadas temporalmente en la frontera definida por dicho radio crítico. Dejando a un lado estas y otras muchas especulaciones teóricas, la astronomía moderna está dedicando un gran esfuerzo al descubrimiento y estudio de agujeros negros con diferentes masas. Aunque llegados a este punto, surge una pregunta evidente: ¿cómo

encontrar un objeto que no emite luz en el cielo nocturno?, ¿hacia dónde orientar nuestros telescopios?

De forma genérica, un agujero negro es una región del cielo con condiciones físicas extremas. No solamente el campo gravitatorio es extremadamente intenso, haciendo necesario el uso de la teoría de la relatividad general para su descripción [1], sino que también se espera que su velocidad de rotación y su campo magnético sean extraordinariamente grandes. Si imaginamos un objeto aislado con masa estelar que colapsa conservando su momento angular y su masa, a medida que el radio se reduce, la velocidad de rotación debe aumentar para compensar la disminución de tamaño. De forma similar, si se conserva el flujo magnético [2], el colapso conduce a un aumento del campo magnético. Estas ideas generales se confirman cuando se comparan las propiedades de estrellas colapsadas de neutrones (púlsares [3] con una radio de unos 10 kilómetros) y estrellas normales. Las estrellas de neutrones pueden alcanzar velocidades de rotación próximas a 100000 kilómetros por segundo, y albergar campos magnéticos de un billón de Gauss. Por consiguiente, en las inmediaciones de un objeto compacto (estrella de neutrones o agujero negro) podemos encontrar un disco de gas caliente: gas capturado por su enorme campo gravitatorio, distribuido en el plano perpendicular al eje de rotación, y calentado por la fricción entre regiones adyacentes del disco. Dicho gas caliente emitirá radiación térmica. También se puede emitir radiación no térmica, cuando partículas cargadas son aceleradas por campos eléctricos y magnéticos en las cercanías de la estructura colapsada.

13.1 Agujeros negros estelares.

Una estrella con masa solar no puede formar un agujero negro. Tras consumir el hidrógeno y el helio en su corazón mediante procesos de fusión nuclear, la envoltura es arrojada al espacio interestelar (nebulosa planetaria) y aparece un núcleo estable de carbono del tamaño de la Tierra (enana blanca). Se requieren decenas de masas solares para alcanzar temperaturas interiores más altas, continuar las fases de fusión nuclear y colapso [4], y llegar a una configuración final en la cual el colapso gravitatorio no puede ser detenido por ningún mecanismo conocido. Entonces se forma un agujero negro estelar. Algunas estrellas masivas evitan la formación de un agujero negro desprendiéndose de forma violenta de su envoltura (explosión de supernova). Aunque el núcleo residual de neutrones (estrella de neutrones) es una estructura compacta, existe una presión cuántica que evita la formación de un agujero negro. Las estrellas de neutrones tienen un radio (aproximadamente 10 kilómetros) ligeramente mayor que el radio crítico para su masa: unos 6 kilómetros para una masa típica de 2 veces la masa solar. Cálculos detallados indican que esta presión cuántica puede soportar hasta 3 masas solares, y por lo tanto, la masa de un agujero negro estelar debe exceder este valor.

¿Conocemos algún candidato a agujero negro estelar? Si, conocemos varios candidatos que forman parte de sistemas binarios. Un objeto compacto en un sistema binario puede devorar el gas de la envoltura de su estrella compañera. Como ya hemos comentado, se forma un disco de gas caliente, y el material se precipita sobre el objeto colapsado realizando una trayectoria espiral (la fricción elimina momento angular, y el gas adquiere un pequeño movimiento radial de caída). En los anillos más internos del disco se alcanzan temperaturas de millones o decenas de millones de grados, y por consiguiente, el entorno del objeto compacto es detectable mediante un telescopio

espacial de rayos X. Tras detectar una binaria compacta de rayos X, la dinámica de la compañera puede entonces usarse para estimar la masa de la estructura colapsada, y así, encontrar un candidato a agujero negro cuando esta exceda las 3 masas solares.

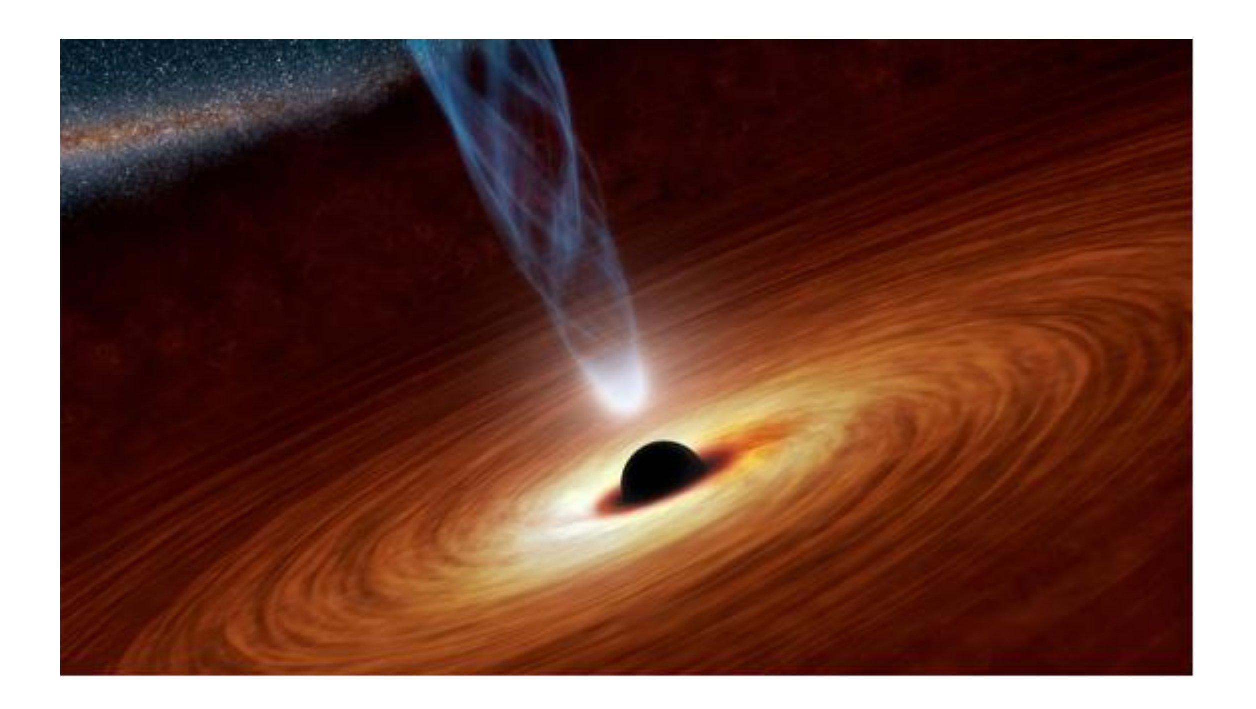
Entre los candidatos más destacados dentro de la Vía Láctea (la galaxia donde reside el Sol y la Tierra) está Cyg X-1 en la constelación del Cisne. Fue descubierto en el año 1964 y es una de la fuentes cósmicas mas intensas de rayos X. El objeto compacto tiene una masa de 14,8 veces la masa solar, lo que implica un radio crítico de unos 44 kilómetros. Otros candidatos interesantes son: GRO J0422+32 y GW150914. El primero (GRO J0422+32) fue descubierto en 1992 en una binaria de rayos X. El objeto compacto en dicha binaria es el candidato menos masivo a agujero negro estelar, ya que solo tiene 4 veces la masa del Sol. Su proyección en un plano tiene un área algo mayor que la superficie de la isla de La Gomera. El candidato GW150914 es uno de los hallazgos más excitantes del siglo actual. No fue detectado en rayos X, sino mediante observaciones de ondas gravitatorias [5]. En Febrero de 2016, se anunció la observación de una colisión entre dos agujeros negros de 36 y 29 masas solares en una galaxia situada a más de mil millones de años luz de la Tierra. Un sistema binario puede estar formado por dos objetos compactos, y a medida que transcurre el tiempo, pierde energía rotacional mediante la emisión de ondas gravitatorias. Finalmente, ambos objetos colapsados colisionan para formar una estructura de mayor masa. En el proceso de colisión y fusión de GW150914 se piensa que se ha formado un agujero negro de más de 60 masas solares con el radio típico de un asteroide.

13.2 Agujeros negros en núcleos galácticos.

Actualmente se piensa que existe un agujero negro muy masivo en el centro de casi todas las galaxias. En los núcleos galácticos, se espera una densidad inicial de estrellas y gas muy elevada, y en dicho ambiente, es razonable imaginar un gran colapso al cabo de cierto tiempo. Las estrellas en la región central de la Vía Láctea describen órbitas elípticas que son consistentes con la acción gravitatoria de un objeto compacto con una masa de unos 4 millones de veces la masa solar. El radio critico (tamaño radial de un hipotético agujero negro en el centro de la Vía Láctea) para esta masa es de 0,08 unidades astronómicas, siendo 1 unidad astronómica la distancia entre el Sol y la Tierra. Sin embargo, las estrellas observadas no se aproximan tanto al monstruo masivo, y datos recientes indican que el radio de la estructura central debe ser inferior a 45 unidades astronómicas (distancia de máxima aproximación de estrellas), o equivalentemente, a unos 560 radios críticos. ¿Se trata de un agujero negro? Muy probablemente si… Aunque las estrellas no se acercan al radio crítico en su movimiento, 45 unidades astronómicas es la distancia entre el Sol y el planeta menor Makemake. Es decir, tenemos varios millones de veces la masa del Sol dentro de un radio menor que el del sistema solar, y no se conoce ningún mecanismo físico capaz de evitar el colapso de semejante estructura.

Aparte de las galaxias normales como la Vía Láctea, existen galaxias con un núcleo activo. Un tipo de especial relevancia está constituido por los llamados cuásares [3]. Los cuásares son núcleos activos de galaxias lejanas que muy probablemente albergan un agujero negro con una masa típica entre cien y mil millones de veces la masa del Sol, dentro de una región con un tamaño menor que el del sistema solar. Las observaciones de cuásares indican que el hipotético agujero negro central está rodeado por un disco de gas caliente, desde donde se emite la luz visible y ultravioleta. Se piensa

que los rayos X provienen de una corona muy caliente sobre dicho objeto central, y en aproximadamente el 10% de los casos, las observaciones también sugieren la presencia de un chorro de partículas energéticas en una dirección perpendiculares al disco y emitiendo radiación no térmica (ver figura adjunta; origen: NASA/JPL – Caltech).



Los anillos más internos del disco emitiendo radiación térmica estarían situados a unos pocos radios críticos del agujero negro, de modo que tales anillos tan lejanos y diminutos (tamaño de un sistema estelar) no pueden resolverse espacialmente de una forma directa. Sin embargo, actualmente se usan algunos métodos indirectos para estudiar la geometría interna del disco, en las proximidades del supuesto monstruo oscuro. Los resultados indican que cientos o miles de millones de masas solares deben estar confinadas en zonas con tamaños radiales de unos cuantos radios críticos, apoyando la presencia real de agujeros negros muy masivos en cuásares.

Notas:
1 Ver capítulo 37.
2 El flujo magnético es el producto del campo magnético por la superficie.
3 Ver capítulo 12.
4 Cada vez que un material interno se agota y las correspondientes reacciones nucleares cesan, se forma una estructura inactiva compuesta principalmente por el producto de la fusión de dicho material. Por ejemplo, cuando se consume el corazón de hidrógeno, se forma un núcleo inactivo dominado por helio. La temperatura no es suficiente para comenzar la fusión del producto, y no se genera una presión que compense la atracción gravitatoria entre las diversas partes del corazón estelar. Entonces el corazón colapsa y se calienta hasta que se alcanza la temperatura umbral para la fusión del producto. El núcleo estelar vuelve a estar activo, y la nueva presión compensa total o parcialmente la contracción gravitatoria del mismo.
5 Las ondas gravitatorias son el equivalente gravitatorio de las ondas electromagnéticas (a las que hemos llamado radiación, o también luz). Así como las cargas eléctricas aceleradas emiten ondas electromagnéticas, cuando existen masas aceleradas por efectos gravitatorios se pueden emitir ondas gravitatorias (ver también el capítulo 14).

Bibliografía complementaria:
"Historia del Tiempo: del Big-Bang a los Agujeros Negros", Stephen W. Hawking.
"La Evolución de Nuestro Universo", Malcolm S. Longair.
"Agujeros Negros: 2^4 preguntas", IFCA (CSIC-UC) – https://ifca.unican.es/es-es/Documents/Teaching-Materials/Agujeros_negros.pdf.

Luis Julián Goicoechea Santamaría
Doctor en Ciencias Físicas
Profesor de Astronomía y Astrofísica – Universidad de Cantabria

Capítulo 14
¿Por qué ha tenido tanta repercusión el descubrimiento de las ondas gravitacionales?
(Por Miguel Zumalacárregui Pérez)

En noviembre de 1915, Albert Einstein daba el toque final a lo que sería una de las joyas del conocimiento humano. Tras una década de trabajo consiguió sintetizar las ecuaciones que gobiernan la teoría de la relatividad general, una construcción matemática de gran belleza que describe la gravedad en términos de la curvatura del espacio tiempo. Einstein dedicó diez años a completar esta gesta, pero la aventura no hacia sino empezar, pues en las próximas décadas se fueron entendiendo y confirmando las predicciones más psicodélicas de esta hermosa teoría, desde la expansión del universo, indicando que existe un origen del tiempo, hasta la existencia de agujeros negros en los que el tiempo se detiene. Cien años después solo una predicción clave de la relatividad general quedaba por confirmarse: la existencia de ondas gravitacionales, ínfimas distorsiones del tejido espaciotemporal que viajan a la velocidad de la luz. Este fenómeno pudo demostrarse por fin, un siglo después de haberse formulado la teoría.

La detección de ondas gravitacionales es y será uno de los hallazgos científicos más importantes del siglo XXI. Se trata de un logro tecnológico sin precedentes, fruto de décadas de esfuerzo y una intensa colaboración a nivel internacional. Este hito abre una nueva ventana al universo que nos permitirá escuchar los fenómenos más extremos del cosmos, fenómenos que de otra forma permanecerían ocultos y que podremos usar para entender la gravedad en mayor profundidad. Y esto es esencial, ya que la gravedad está íntimamente ligada a muchos de los grandes problemas en física teórica, desde la unificación de las fuerzas hasta el origen del universo. Independientemente de como lleguen a resolverse estos misterios, las ondas gravitacionales nos ayudarán a entenderlos y nos abrirán la puerta a fenómenos nuevos e inesperados.

Las ondas gravitacionales no son muy distintas de las ondas que se producen en la superficie de un estanque, las ondas que propagan el sonido en el aire, o las ondas electromagnéticas que dan lugar a la luz visible o la radio: se generan por una fuente, se propagan hacia el exterior de ésta, perdiendo intensidad en el camino, y producen ciertos efectos a su paso que permiten detectarlas con las técnicas adecuadas. Una diferencia entre las ondas gravitacionales y electromagnéticas es que las primeras se parecen más al sonido: son emitidas por la totalidad de un sistema y no solo por su superficie. Por este motivo no se habla de "ver" ondas gravitacionales, sino que se las "escucha". Y hay que escucharlas con suma atención, pues la gravedad es una fuerza extremadamente débil. Sabemos esto porque con nuestras piernas somos capaces de vencer la fuerza que ejercen las casi 6 *cuatrillones* de toneladas de la tierra ($6 \cdot 10^{24}$ kg) [1]. Esta debilidad convierte el estudio de las ondas gravitacionales en una ciencia tremendamente desafiante, pues incluso los fenómenos más violentos del universo producen distorsiones ínfimas y detectarlas desafía los límites de nuestra tecnología.

La primera onda gravitacional jamás detectada, GW150914, cruzó la tierra el 14 de septiembre de 2015. Sin embargo su viaje empezó hace más de mil millones de años, cuando dos agujeros negros que estaban orbitando en una galaxia lejana colapsaron y se fundieron en un agujero negro mayor. Aunque cada uno de estos agujeros negros tenía una masa unas 30 veces mayor que nuestro Sol (¡y este es unas cien millones de veces más pesado que la Tierra!) la fusión se produjo en menos de dos décimas de segundo. Este proceso, tan rápido y violento liberó una cantidad de energía correspondiente a 3 veces la masa del Sol en forma de ondas gravitacionales. Al igual que el sonido del trueno es apenas audible cuando la tormenta es distante, el rugido de este proceso monstruoso apenas es perceptible a su llegada a la tierra.

Detectar las ondas gravitacionales ha sido un desafío épico a nivel tecnológico. Una onda gravitacional tiene el efecto de acortar el espacio en una dirección y estirarlo en otra. Pero esta onda se atenúa y después de un viaje de mil millones de años, incluso la colisión de dos agujeros negros mucho más masivos que el Sol tiene un efecto imperceptible. La distorsión del espacio es por apenas un factor 10^{-21}, es decir, una milésima de una trillonésima. Podemos entender lo minúsculo de este factor con una analogía. Imaginemos que en lugar de medir distorsiones del espacio, nos interesáramos por fluctuaciones monetarias, por ejemplo nuestros ahorros en el banco o el valor de una acción en la bolsa. En ese caso, una fluctuación de 10^{-21} es equivalente a modificar el valor de toda la economía mundial en una diezmillonésima de euro. Tal cantidad es tan ínfima que está muy por debajo de la menor fracción de divisa que se considera en ningún mercado.

La colaboración LIGO (Laser Interferometer Gravitational-Wave Observatory) ha sido capaz de efectuar esta medida histórica usando un sistema extremadamente sofisticado para detectar estas perturbaciones en el espacio-tiempo. Cada uno de los detectores de LIGO consiste en dos rayos láser perpendiculares, cuya interferencia permite medir cambios en la distancia relativa que recorre cada uno de los rayos. La pequeñez de las ondas gravitacionales hace que el estiramiento de estos brazos es menor que el radio de un protón. Para sobreponerse a la pequeñez de esta distorsión los espejos que reflejan los láser están suspendidos en un sistema cuádruple de péndulos, que los aísla del movimiento de la tierra. A pesar de todos estos trucos, multitud de efectos (que incluyen pequeños terremotos, coches pasando a kilómetros del detector o la vibración térmica de los espejos) son mucho mas fuertes que la señal, por lo que LIGO cuenta con dos detectores situados a 3000km en extremos opuestos de EEUU. Al contrario que todos estos efectos, una onda gravitacional produce un patrón distintivo en ambos detectores: la misma señal es recogida en ambos puntos, pero con unos milisegundos de retraso que la onda gravitacional tarda en recorrer la distancia que los separa, viajando a la velocidad de la luz.

El patrón de ondas gravitacionales detectado por LIGO es justo lo que se espera escuchar en las últimas órbitas de dos agujeros negros con una masa unas 30 veces mayor que la de nuestro Sol, su fusión en un agujero negro con la masa total y la relajación de éste hasta un estado estable. Sabemos esto por la frecuencia a la que se observan las distorsiones en los láser de LIGO y su variación temporal, ya que la onda oscila más rápido a medida que ambos agujeros negros se acercan. Los detalles técnicos son muy complejos, pero la idea se puede entender con una analogía musical. Sabemos distinguir la misma melodía tocada en distintos instrumentos, ya que, incluso tocando las mismas notas, un contrabajo, un órgano y un trombón son completamente diferentes, y un experto puede incluso distinguir entre ligeras variaciones en el mismo tipo de

instrumento. Un punto importante para saber que son agujeros negros (y no otro objeto muy compacto) es por la velocidad de las órbitas: si los objetos fueran más grandes que un agujero negro tendrían que desplazarse más rápido que la luz para reproducir las frecuencias que observamos, lo cual contradice la teoría de Einstein.

Vale la pena detenerse en este punto: LIGO no solo ha detectado ondas gravitacionales, sino que también ha sido capaz de observar dos agujeros negros orbitando y fundiéndose en un agujero negro mayor. Este evento abre una nueva ventana al universo y a los fenómenos más violentos que ocurren en él. Los agujeros negros son los objetos más extremos conocidos, el estado final en la evolución de estrellas tan densas que la gravedad las aísla del resto del universo e impide hasta a la misma luz escapar de su interior (ver Capítulo 13). Es muy difícil estudiar estos objetos, ya que por definición no emiten ninguna señal. Pero ahora todo esto puede cambiar, pues hemos ganado un sentido con el que estudiar los agujeros negros, así como otros fenómenos gravitatorios. Ya no solo podemos ver el universo, ahora también podemos escucharlo.

Las pocas pero espectaculares observaciones de ondas gravitacionales que hay hasta la fecha no hacen más que arañar la superficie de lo que puede llegar a ser esta nueva ciencia. Cada nuevo canal de observación al que se ha abierto la astronomía ha dado lugar a descubrimientos sorprendentes e interesantes. Un claro ejemplo es la radioastronomía, que estudia objetos astronómicos a través de las ondas de radio, como las que empleamos para la comunicación con dispositivos móviles. La apertura de los cielos a observaciones a través de ondas de radio dio lugar al descubrimiento de los quasares, objetos tan brillantes y tan distantes que solo pueden ser causados por procesos extremos en galaxias muy lejanas. Se cree que los quasares son la radiación emitida por estrellas y nubes de gas al ser devoradas por agujeros negros inmensos que habitan en el centro de las galaxias, cada uno millones de veces más masivo que nuestro Sol (ver Capítulo 12). Las ondas gravitacionales nos proporcionarán nuevas pistas sobre los agujeros negros. Pero más allá de observar estos misteriosos seres, o de ganar nueva información sobre otros objetos como las estrellas de neutrones, sin duda lo más vertiginoso es la posibilidad de observar objetos hasta ahora desconocidos.

Más allá de observar fenómenos y objetos en el universo distante, las ondas gravitacionales nos permitirán entender mejor las leyes fundamentales que rigen nuestro universo a nivel microscópico. Y en concreto podremos estudiar propiedades de la que es con diferencia la fuerza peor entendida de la naturaleza: la gravedad. A pesar de ser uno de los fenómenos más cotidianos, de agarrarnos cada minuto de nuestra existencia, la gravedad es sin duda el fenómeno más misterioso de la física fundamental. Una cantidad abrumadora, una fracción mayúscula de los problemas abiertos en física fundamental están conectados de alguna manera a los fenómenos gravitatorios y la naturaleza del espacio-tiempo. La gravedad es parte integral en el entendimiento de problemas que van desde el origen del universo y su destino hasta la descripción microscópica de la gravedad (a través de su unificación con la mecánica cuántica), pasando por la naturaleza de los agujeros negros.

Hablamos de una ciencia que está en su infancia, y por el momento es difícil de imaginar en qué forma concreta el estudio de las ondas gravitacionales nos podrá ayudar a entender estos problemas tan profundos. Aunque es una idea especulativa, es muy posible que las ondas gravitacionales nos ayuden a entender el mecanismo por el que el universo se expande de manera acelerada (ver Capítulo 52). Este efecto contradice la

naturaleza atractiva de la gravedad predicha por la teoría de Einstein, creando así uno de los grandes misterios de la ciencia contemporánea. Para explicarlo se han propuesto muchos modelos basados en teorías de gravedad alternativas. Éstas permiten que la gravedad sea más débil a escalas cosmológicas, permitiendo que las galaxias se alejen cada vez más rápido unas de otras. Pero muchas de estas teorías también predicen modificaciones substanciales en el comportamiento de las ondas gravitacionales. Por ejemplo, una clase muy importante de modelos predicen que estas viajan a velocidades superiores a la luz, haciendo que las señales de eventos lejanos se observen con un lapso entre la señal gravitatoria y su análogo electromagnético. Este test no puede hacerse con agujeros negros, ya que no emiten luz al fusionarse, pero cuando seamos capaces de observar las ondas gravitacionales de objetos luminosos podremos descartar muchos modelos de energía oscura.

La gravedad es única y hemos ganado un sentido con el que poder escucharla. Es la fuerza que da forma al universo, ensambla las galaxias y las agrupa en estructuras aun mayores. La gravedad aparece íntimamente ligada a muchos de los problemas no resueltos por la física contemporánea [2], algunos tan antiguos como el origen del universo. También está inseparablemente ligada a la naturaleza misma del espacio-tiempo, nos amarra a nuestro planeta y nos limita a viajar a velocidades que son ridículas en la escala de las galaxias. Como el amor, la gravedad es un fenómeno atractivo y universal, a menudo demasiado obvio, débil o inalcanzable como para llamar nuestra atención. Y como esta otra fuerza irresistible, nunca dejará de sorprendernos.

Notas:
[1] Para una introducción a la notación en potencias de 10 véase www.fronterad.com/?q=pequeno-mapa-mundo-en-potencias-diez-i-contando-cifras
[2]Problemas no resueltos en física: https://en.wikipedia.org/wiki/List_of_unsolved_problems_in_physics

Bibliografía:
Nota de prensa de la colaboración LIGO https://www.ligo.caltech.edu/detection
"The Perfect Theory: A Century of Geniuses and the Battle Over General Relativity", Pedro G. Ferreira.
Gravitational Waves Explained, Jorge Chan https://www.youtube.com/watch?v=4GbWfNHtHRg

Miguel Zumalacárregui Pérez
Doctor en Física
Investigador Marie Curie - University of California at Berkeley
Nordita Fellow – Nordic Institute for Theoretical Physics

Capítulo 15
¿Hasta dónde puede llegar el proceso de fusión dentro de una estrella?
(Por Ricardo Dorda Laforet)

De forma simplista puede describirse a las estrellas como bolas de gas incandescente, pero aunque visual, es una aproximación que omite el elemento fundamental que rige su existencia: el equilibrio. Las estrellas son el producto de dos fuerzas titánicas enfrentadas entre si. Por un lado está su campo gravitatorio, causado por su propia masa y que tiende a comprimirla. Por otro lado, la energía que se genera en su núcleo mediante fusión nuclear y que produce una presión interna que tiende a expandirla [1], igual que ocurre en una explosión. Así, toda la vida de la estrella, con sus diferentes fases evolutivas desde su nacimiento hasta su muerte, vendrá determinada por la contraposición de ambas fuerzas.

15.1 Nacimiento

Cuando las grandes nubes interestelares de gas y polvo colapsan sobre si mismas [2], forman grumos de mayor densidad en su interior (llamadas protoestrellas). Tales puntos se convierten en focos del colapso, atrayendo hacia sí toda la materia cercana, lo que aumenta progresivamente su densidad. En cierto momento, la presión, densidad y temperatura son tan altas en el centro de esos grumos, debido al peso de toda la materia que cae hacia allí, que en su centro se inician las reacciones de fusión nuclear.

La fusión nuclear consiste en la unión de núcleos atómicos ligeros en otros más pesados. Sin embargo, para que dos núcleos atómicos (cuya carga eléctrica es netamente positiva) lleguen a fusionarse, poniéndolo en términos muy simplificados, hace falta que impacten entre ellos con una energía (velocidad) tremenda, suficientemente alta como para que venzan su repulsión natural (debido a que ambas son cargas eléctricas del mismo signo). Pese al coste energético que requiere el impacto, una vez se fusionan en un único núcleo alcanzan un estado muchísimo más estable que el que tenían individualmente, por lo que se libera una enorme cantidad de energía. Una forma de visualizar esto es pensar, por ejemplo, en dos palos de madera. Si frotamos los palos entre ellos despacio no conseguiremos nada: no estaremos dándoles suficiente energía como para que se prendan fuego. Sin embargo, si frotamos con suficiente velocidad, es decir, les damos suficiente energía inicial, los palos se calentarán tanto que empezarán a arder y su combustión nos proporcionará mucha más energía de la que nos costó que prendieran.

El gas a partir del que nacen las estrellas consiste mayoritariamente en Hidrógeno (~75% de su masa) y Helio (~25%) con una pequeña fracción de otros elementos más pesados. Al irse incrementando la presión y la temperatura en el núcleo estelar debido a la materia que va cayendo hacia él, lo primero que comenzará a

fusionarse son núcleos del elemento más ligero (el Hidrógeno), pues la repulsión entre ellos es menor que entre núcleos mayores (como el Helio) cuya carga positiva es mucho mayor y, por tanto, que se fusionen requieren mucha más energía. Cuando empieza a producirse energía en el núcleo gracias a la fusión, ésta tiende a expandir la estrella, lo que detiene el colapso de la estrella. Eso detiene súbitamente el incremento de densidad y temperatura en el núcleo estelar, pues no sigue comprimiéndose y por eso no llegan a alcanzarse las temperaturas que requieren los núcleos más pesados para fusionarse.

Dependiendo de la masa total de la protoestrella, las condiciones de su núcleo serán muy diferentes. A mayor masa, tendrá una mayor presión interna que le permitirá mantener un núcleo de mayor tamaño y a mayor temperatura. Uno podría plantearse que un núcleo mayor implica una mayor cantidad de Hidrógeno en él, por lo que una estrella de más masa tardaría mucho más en agotar su núcleo que una menos masiva. Sin embargo, esto es justo al revés.

Existen dos mecanismos para la fusión del Hidrógeno, la cadena protón-protón y el ciclo CNO, y ambos se dan en todas las estrellas sin importar su masa. Ahora bien, de ambos, el que requiere menos energía para activarse es la cadena protón-protón. Sin embargo, su eficiencia depende poco de la temperatura si lo comparamos con el ciclo CNO, que es mucho más sensible a la temperatura, de forma que es mucho más eficiente cuanto mayor sea ésta. Así, ocurre que en las estrellas con una masa 1.3 veces la del Sol o menor el mecanismo dominante es la cadena protón-protón. Sin embargo, en las estrellas de más masa la temperatura es suficientemente alta como para que domine el ciclo CNO, lo que acelera enormemente el consumo de Hidrógeno y, por tanto, la producción de energía. La consecuencia de esto es que una estrella más masiva consumirá el Hidrógeno en su núcleo mucho más rápidamente que una estrella menos masiva. Tanto que, pese a que las reservas de Hidrógeno en el núcleo son mucho más grandes para la estrella más masiva, ésta las agotará muchísimo antes que la estrella menos masiva. Así, una estrella como el Sol tarda unos 10.000 millones de años en quemar el Hidrógeno de su núcleo, pero una estrella con 30 veces la masa del Sol solo tarda unos pocos millones de años. Así mismo, las estrellas menos masivas que el Sol tardan en agotar sus núcleos muchísimo más que éste.

En consecuencia, las estrellas más masivas emiten mucha más energía por segundo (son mucho más luminosas) que las menos masivas, lo que hace que sean más grandes y que su superficie esté mucho más caliente [3].

15.2 Vida

La etapa de la vida de las estrellas durante la que fusionan Hidrógeno en sus núcleos se llama fase de Secuencia Principal, y es la de mayor duración. La fusión de Hidrógeno produce Helio, que se va acumulando en el núcleo ya que su temperatura no es suficientemente alta como para que pueda fusionarse. Cuando el Hidrógeno se agota en el núcleo, deja de producirse la energía que contenía la compresión gravitatoria de la estrella. Sin la fusión nuclear, nada evita que el interior de la estrella empiece a colapsar hacia el centro y, al tiempo que esto ocurre en el interior de la estrella, sus capas externas tienden a expandirse de forma espectacular. La estrella se convierte así en una gigante (o supergigantes) roja [3]. Ahora bien, en este punto depende de la masa estelar qué ocurre a continuación.

Si la masa es relativamente baja [3], de 2.25 veces la del Sol (masas solares) o menor, el núcleo se comprime hasta que se degenera. Es decir, que su presión interna deja de ser principalmente de origen térmico para ser de origen cuántico. Esto, planteado de forma simplista, significa que la materia está tan compacta que los núcleos atómicos se agolpan entre ellos sin poder invadir el espacio de los demás. Dado que la materia del núcleo degenerado ya no puede "ocupar menos espacio", absorbe la compresión debida al colapso gravitatorio de las capas superiores en forma de temperatura (con los átomos vibrando cada vez más rápido en un espacio absolutamente restringido). Debido a lo compacto que están los núcleos atómicos en ese estado, el núcleo estelar tiene una altísima conductividad, es decir, que todo el núcleo incrementa su temperatura prácticamente al mismo tiempo pese a su enorme tamaño.

Las estrellas de baja masa, el 23% de la del Sol o menos, carecen de la masa necesaria como para comprimir su núcleo hasta el punto en que su temperatura permita la fusión del Helio, muriendo como estrellas enanas blancas de Helio. Sin embargo, las estrellas con masas mayores (pero menores a 2.25 masas solares) sí alcanzan una temperatura en su núcleo tan alta que el Helio comienza a fusionarse (a unos 100 o 200 millones de Kelvin). La cuestión es que como todo el núcleo llega a la temperatura necesaria casi a la vez, todo el Helio se fusiona simultáneamente en Carbono, lo que libera repentinamente una enorme cantidad de energía que vuelve a expandir el núcleo, terminando la degeneración. A este evento se lo conoce como Flash de Helio. Después de eso, la estrella es capaz de seguir fusionando Helio en una capa alrededor de su núcleo inerte de Carbono (que se comprime hasta el punto en que se degenera, ya que no produce energía propia). Además, envolviendo a la capa de fusión de Helio hay también una capa a suficiente temperatura como para que el Hidrógeno en ella se fusione en Helio, el cual va cayendo hacia la capa inferior, dónde se fusiona en Carbono. Cuando la estrella agota el Hidrógeno y el Helio disponibles en su núcleo, muere como una estrella enana blanca, pues carece de la presión suficiente como para que el Carbono degenerado de su núcleo pueda fusionarse en elementos aún más pesados.

Volvemos ahora a las estrellas con masas superiores a 2.25 masas solares. Estas estrellas poseen suficiente masa como para que sus núcleos alcancen las altísimas temperaturas necesarias para la fusión del Helio sin pasar por una fase de materia degenerada. Además, alrededor del núcleo de Helio hay una capa de fusión de Hidrógeno, que alimenta con nuevo Helio al núcleo. Entre tanto, en el centro del núcleo se va acumulando, Carbono y Oxígeno, productos de la fusión del Helio. Si la estrella tiene menos de 8 masas solares, el núcleo de Carbono y Oxígeno nunca alcanzará la temperatura que permita su fusión y la estrella termina muriendo cuando agota su Hidrógeno y su Helio disponibles, dejando tras de sí una enana blanca.

Por el contrario, en las estrellas con una masa superior a 8 veces la solar, el núcleo alcanza temperaturas que permiten que Carbono y Oxígeno comiencen a fusionarse. Así, se forma una capa de fusión de estos elementos alrededor del centro del núcleo, dónde se acumula el producto de su fusión: Neón y Magnesio. Como en estas estrellas las temperaturas del núcleo son suficientemente altas, estos elementos también empezarán a fusionarse en una capa alrededor del producto de su fusión, y así sucesivamente, de forma que el interior de la estrella se estructura en una serie de capas de fusión, como las capas de una cebolla. El producto de la fusión de cada capa va cayendo a la siguiente capa más interior, dónde se fusiona a su vez en elementos más pesados que continúan su camino, capa a capa, hasta el centro de la estrella.

15.3 Muerte

Como hemos visto, la muerte para las estrellas de 8 masas solares o menos ocurre cuando en su núcleo se agotan los elementos que puede fusionar, quedando solo aquellos demasiado pesados en un estado degenerado. Esto da lugar a una estrella enana blanca, que no es otra cosa que el antiguo núcleo de la estrella, ahora inerte pero tremendamente caliente (miles de veces más que la superficie una estrella viva). Mientras, los restos de las capas más superficiales de la estrella son eyectados, dispersándose hacia el medio interestelar y formando en el proceso lo que se conoce como Nebulosa Planetaria.

Por otro lado, las estrellas con más de 8 masas solares tienen núcleos capaces de fusionar elementos cada vez más pesados... hasta que empieza a producirse Hierro-56. Este elemento es el más estable de la tabla periódica. Es decir, que fusionarlo requiere mucha más energía de la que libera. Así, el Hierro-56 se va acumulando rápidamente en el centro de la estrella, completamente inerte. Al principio, su degeneración evita su colapso, pero la estrella continúa acumulando Hierro hasta que su masa es demasiado grande como para que la degeneración atómica pueda sostener su peso (es decir, se alcanza la masa de Chandrasekhar, que es igual a 1.4 masas solares). En ese punto, ningún mecanismo puede evitar el colapso del Hierro, que colapsa a velocidades relativistas, lo que provoca que se libere una cantidad enorme de radiación gamma que destruye los núcleos de Hierro, dejando tras de sí una estrella de neutrones o un agujero negro (según cuán masiva fuera la estrella) y provocando una repentina onda expansiva hacia fuera. Ésta causa la compresión súbita de muchas de las capas internas de la estrella tan enorme, que su materia se fusiona simultánea e instantáneamente. A este proceso se lo conoce como Supernova tipo II, lo que marca la muerte de la estrella de forma absolutamente espectacular, ya en una fracción de segundo libera la misma energía que millones de soles durante toda su existencia. Y es precisamente en esta tremenda explosión, que tanta energía libera, dónde se generan todo tipo de elementos más pesados que el Hierro-56 (desde el Cobalto hasta el Plutonio) y que nunca podrían haber nacido por fusión en el núcleo de ninguna estrella.

Notas:

1. En aras de la precisión hay que advertir al lector que esto es cierto solo para las estrellas “vivas”, es decir, que producen su propia energía. En contraposición hay tipos de estrellas que pueden describirse como “muertas”, restos de estrellas convencionales que ya no producen energía por fusión en sus núcleos, como las enanas blancas y las estrellas de neutrones.
2. Ver capítulo 11.
3. Ver capítulo 17.

Ricardo Dorda Laforet
Doctor en Astrofísica
Investigador post-doctoral en la Universidad de Alicante

Capítulo 16
¿Qué es el viento solar?
(Por Antonio Guerrero Ortega)

El sistema solar, en el que orbitan los planetas más cercanos, está definido por un entorno dominado por el viento solar. En el inmenso espacio de la Vía Láctea (nuestra galaxia), compuesta de millones de estrellas, el sistema solar sostenido por la presencia del viento solar que emana de la atmósfera del Sol es como una pequeña burbuja que envuelve a nuestra estrella en su viaje por la galaxia. Esta burbuja dominada por el viento solar es lo que llamamos Heliosfera.

La primera sospecha de que existía una especie de viento que soplaba desde el Sol en todas direcciones existe desde que se observan los cometas. Los cometas son objetos que viajan, en parte o totalmente, por el entorno de la Heliosfera (medio interplanetario) y que la humanidad ha podido observar fácilmente desde la Tierra desde mucho antes de la era espacial. La característica más llamativa de ellos es lo que comúnmente llamamos la cola del cometa, el rastro que parecen dejar en su viaje espacial debido a que están compuestos de materia que se desprende cuando se acercan al Sol. Lo que hizo pensar por muchas generaciones que tenía que existir un viento que sople alejándose del Sol en todas direcciones era el hecho de que la cola de los cometas se aleja siempre del Sol en la dirección radial, independientemente de si el cometa se dirige hacia el Sol o se aleja del mismo. Lo lógico sería pensar que la cola del cometa son partículas que se quedan detrás en su viaje por el medio interplanetario, pero no es así, pues cuando el cometa se aleja del Sol, la cola se sitúa en la parte de delante, dando a entender que existía una fuerza con origen en el Sol que las hacía desprenderse en esa dirección. Una de las preguntas frecuentes que se hacían era si ese viento podría ser debido a la presión de la luz solar (la misma que ilumina la Tierra) o si se necesitaría algún otro tipo de partículas para provocar ese efecto de arrastre de las partículas de la cola del cometa. Además, algunos cometas muestran una doble cola, una cola blanca brillante en la dirección radial y otra cola de otro color, de luz más débil y dirigida en una dirección distinta. Como veremos más adelante, muchos de los efectos observados en la época anterior a la era espacial, solo pueden ser explicados por la presencia de las partículas con masa en un estado de plasma. Por ejemplo, la conservación de la estructura de campo magnético solar a distancias muy lejanas de su superficie hace posible la captura de partículas ionizadas de la cola de algunos cometas, dirigiéndolas en la dirección de las líneas de campo y emitiendo luz con el color característico de dichos iones.

16.1 Radiación solar.

El Sol emite dos tipos muy diferentes de radiación, una con masa y otra sin masa. La radiación sin masa es luz (una onda electromagnética), la cual ilumina y

calienta la Tierra; la radiación con masa son partículas que escapan de la atmósfera solar. La luz tarda en realizar la distancia Sol-Tierra unos ocho minutos debido a que viaja a la máxima velocidad posible en el universo (la velocidad de la luz) gracias a que no posee masa y a que viaja en un medio que no interfiere en su propagación (cuasi-vacío). El otro tipo de radiación corpuscular (partículas con masa) es precisamente la que provoca que el espacio Sol-Tierra no sea un vacío absoluto.

16.2 El estado de plasma.

Para ser capaces de entender la estructura del viento solar y su comportamiento hay que saber que prácticamente todo lo que existe fuera de nuestro planeta está en estado de plasma. El plasma es un estado más como lo son el estado sólido, el estado líquido o el estado gaseoso. En ese orden, el plasma estaría a continuación del estado gaseoso. El orden lo marca el grado energético que poseen las partículas que componen la materia. En el estado sólido, las partículas poseen poca energía, y por tanto las velocidades a las que se mueven unas partículas respecto de las otras es prácticamente nulo. En el estado líquido, las partículas poseen mayor energía y es posible que existan movimientos relativos entre las mismas. En el estado gaseoso, las partículas se mueven a gran velocidad en cualquier dirección. El cuarto estado es el plasma, en el que las partículas poseen tanta energía, que éstas son capaces de separarse en partes más pequeñas. Las partículas más pequeñas que constituyen la materia es lo que se conoce como átomos, aunque sabemos que podemos separar los átomos en sus constituyentes más pequeños, por ejemplo, en protones y electrones. La diferencia estriba en que estos constituyentes poseen la propiedad de carga eléctrica, positiva para los protones y negativa para los electrones y por tanto para separarlos es necesario superar la fuerza electromagnética que los mantiene unidos a esas distancias tan cortas. En el estado de plasma, las partículas poseen tanta energía que sus electrones superan esas fuerzas, quedando en un estado libre, sin pertenecer a un átomo en concreto y en consecuencia creando campos electromagnéticos en el entorno externo. En esta descripción del estado de la materia se tienen en cuenta los movimientos internos relativos de las partículas y no los movimientos por otras fuerzas externas que puedan tener el conjunto de partículas. Así la Tierra puede estar moviendo continuamente a velocidad constante materia en estado sólido, líquido, gaseoso e incluso de plasma.

16.3 ¿Es posible que el Sol emita partículas constantemente?

La gravedad del Sol es mucho mayor que la existente en la Tierra, por tanto debería de ser suficiente para retener todas las partículas de su atmósfera e impedir que se expulse un viento continuo hacia el exterior. Pero no es así, la capa externa de la atmósfera solar (la corona) se acelera en la dirección radial de forma que emite un flujo continuo de partículas hasta los límites del sistema solar. Las causas de porqué sucede de esta forma no es sencilla de entender sin recurrir a ecuaciones de estado y conservación de momento. Lo primero que hay que entender es que la situación se encuentra en estado de equilibrio, dado por las condiciones de contorno, que son las que existen en el origen y el destino del viento solar, es decir, en la corona y en el exterior del sistema solar. El hecho de existir el viento solar como flujo constante de partículas, no quiere decir que no exista equilibrio, simplemente nos dice que el equilibrio no es estático. La clave que hace que la atmósfera solar se escape de su propio dominio se encuentra precisamente en dichas condiciones de contorno; en la temperatura de la

corona, de alrededor de un millón de grados Kelvin, muy superiores a la temperatura de la atmósfera terrestre, y además en la baja presión, densidad y temperatura existente en el exterior del sistema solar (en el medio interestelar).

16.4 Eyecciones de masa coronal y viento solar rápido.

La corona solar, donde se origina el viento solar, se parece muy poco al exterior de nuestro planeta, sobre todo debido a que en su interior tampoco ocurren los mismos fenómenos. El Sol posee una rotación sobre si mismo como la Tierra, pero su superficie no rota como si fuera un sólido, sino como plasma. Su ecuador gira más rápido que los polos y esta circunstancia provoca que el plasma constituya estructuras complejas en las que se producen bucles y regiones con mayor actividad que otras. Esas regiones activas, se ven como zonas brillantes y contienen campos magnéticos muy intensos que mantienen al plasma encerrado en su interior y que impiden que estos sean fuente de viento solar. El estado de estas regiones es en determinadas ocasiones inestable de manera que es posible la reconfiguración explosiva de los campos magnéticos y la consecuente expulsión de plasma hacia el exterior. Estas expulsiones que ocurren de manera esporádica (por tanto no son las responsables de la formación de viento solar) es lo que conocemos como eyecciones de masa coronal. En la superficie del Sol también podemos observar en determinadas ocasiones otras zonas que en este caso son oscuras, normalmente de mayor tamaño que las regiones activas, llamadas agujeros coronales. Estas zonas al contrario que las anteriores, poseen campo magnético abierto, de manera que permiten al plasma de la corona escapar libremente hacia el exterior y por tanto sí que generan viento solar, pero en este caso viento solar rápido que interacciona con el ya existente.

16.5 Campo magnético interplanetario.

Todas estas estructuras de campo magnético tan diferentes en la superficie del Sol giran conjuntamente con la rotación solar. Aunque las partículas de viento solar son expulsadas en dirección radial, las propiedades del campo magnético de cada zona se conservan en su propagación debido a que en este tipo de plasma el campo magnético queda congelado. Esto se debe a que la energía que poseen las partículas es mayor que la que posee el campo magnético. La rotación solar produce por tanto un patrón de campo magnético en el medio interplanetario que visto desde los polos del eje de rotación forma una espiral. Dicho patrón demuestra la conservación de las estructuras magnéticas de la superficie solar a lo largo de su propagación radial y por tanto permiten que desde la Tierra podamos predecir y medir muchas de estas propiedades antes de que la alcancen.

16.6 Falda de bailarina.

El Sol, además, posee una gran estructura de campo magnético, con líneas de campo salientes en un hemisferio y entrantes en el otro. El límite entre una zona magnética y la otra puede entenderse como una línea en la zona del ecuador solar más o menos ondulada dependiendo del momento del ciclo solar. Este ciclo de actividad solar dura once años y en cada ciclo la polaridad positiva (hacia afuera) o negativa (hacia adentro) se alternan. Como se ha comentado anteriormente, toda la gran estructura

magnética del Sol se conserva en su propagación, y esto hace que la línea más o menos ondulada que separa los dos hemisferios magnéticos en el Sol aparezca en el medio interplanetario como una lámina con forma de falda de bailarina, que divide las zonas de campos magnéticos dirigidos hacia el Sol de los dirigidos hacia el exterior del sistema solar. Dicha lámina es una zona neutra magnéticamente que produce interesantes fenómenos.

16.7 La investigación en meteorología espacial.

El viento solar juega un papel muy importante en la relación Sol-Tierra, pues nos conecta de una manera mucho más cercana a nuestra estrella más cercana, permitiéndonos "tocar" prácticamente su atmósfera con instrumentos a bordo de naves que necesitan alejarse de la Tierra una distancia relativamente pequeña respecto de la distancia al Sol. Además, en este flujo continuo de partículas se propagan y suceden multitud de sucesos que aún no conocemos en detalle. Diferentes tipos de estructuras de plasma se establecen e interaccionan en este entorno, y en muchos casos alcanzan nuestro entorno terrestre causando importantes daños en nuestra vida cotidiana. Actualmente la disciplina que estudia el estado del Sol y su relación con el medio interplanetario y los entornos planetarios, para monitorizar y predecir como son afectados se denomina Meteorología Espacial. Son muchas las preguntas abiertas en este campo de investigación, por ejemplo, aún no se conoce con exactitud el origen del viento solar más lento, la Heliosfera es una gran desconocida en los planos que escapan de la eclíptica, en el que orbitan los planetas. Nuevas misiones espaciales son necesarias para aportar los datos necesarios en esas zonas, así como en zonas cercanas al Sol, para ser capaces de desvelar incógnitas como la forma en que las diferentes estructuras de plasma interaccionan aumentando su daño efectivo en los planetas.

Antonio Guerrero Ortega
Doctor en Investigación Espacial
Investigador Postdoctoral de la Universidad de Alcalá

Capítulo 17
¿Es el Sol una gran estrella?

(Por Ignacio Negueruela Díez)

Más de una vez hemos oído decir que el sol es una "estrella pequeña" o "una estrella cualquiera". Los científicos tienden a administrarnos estas píldoras de humildad como vacuna contra nuestra natural tendencia a pensar que todo lo nuestro es especial y mejor que el resto. Pero ¿qué quieren decir con esto? ¿En qué sentido es el Sol una estrella del montón?

17.1 El tamaño del Sol.

El Sol es una bola incandescente de gas con un radio de 700.000 km y una temperatura superficial de 5.800 K. Se nos pueden ocurrir muchos adjetivos que aplicarle, pero, a nuestra escala, seguro que no serán ni diminuto, ni intrascendente. Pero el hombre no es la medida de todas las cosas. ¿Qué lugar ocupa el Sol en la sociedad de las estrellas?

Durante buena parte de su vida, las estrellas producen energía mediante la combustión del hidrógeno en su núcleo para dar lugar a helio [1]. En esta fase, el tamaño de una estrella depende de su masa. A mayor masa, más grande es la estrella (aunque la dependencia no es lineal). Como la temperatura superficial de la estrella también depende de su masa, en los diagramas que muestran propiedades estelares, las estrellas se concentran en una tira relativamente estrecha que se conoce como secuencia principal. Las estrellas más pequeñas, con masas algo menores de un décimo de la solar, tienen radios también unas diez veces menores. Las estrellas más masivas de la secuencia principal tienen radios entre quince y veinte veces mayores que el Sol. Por tanto, vemos que el Sol no es particularmente grande en la escala estelar. Pero incluso esta posición es engañosa. Una vez agotado el hidrógeno en el núcleo, las estrellas se hinchan y su temperatura disminuye. Entonces se hacen mucho más grandes, convirtiéndose en gigantes rojas [2].

Las gigantes rojas típicamente tienen radios entre 10 y 30 veces mayores que el Sol, pero en algunas fases de su evolución pueden llegar a ser bastante más grandes. Una estrella como el Sol puede llegar a alcanzar un radio unas 200 veces mayor que el que tiene ahora (rozando la órbita de la Tierra) cuando sea una gigante de la rama asintótica, justo antes de morir. Otras estrellas más masivas pueden llegar a tener 500 radios solares durante esta fase. Las estrellas mucho más masivas que el Sol se convierten en supergigantes, alcanzando tamaños de más de 1.000 radios solares. Estas son las estrellas más grandes. No estamos seguros del tamaño máximo que puede alcanzar una supergigante roja, pero creemos que anda entre 1.500 y 2.000 radios solares.

Vemos, entonces, que el Sol es una estrella pequeña ahora, pero será mucho mayor en el futuro. El tamaño de una estrella está determinado por su masa y por su edad. Por consiguiente, cuando los astrónomos comparamos estrellas, la propiedad en la que nos fijamos no es tanto el tamaño físico, sino la masa.

17.2 La masa del Sol.

La masa del Sol es tan enorme que necesito casi un renglón entero para escribirla: 2.000.0000.000.000.000.000.000.000.000.000 kg. Los astrónomos medimos la masa de las estrellas en masas solares. Las estrellas más pequeñas tienen una masa en torno al 8% de la masa solar. Si un cuerpo celeste tiene una masa menor, no pueden producirse las reacciones nucleares que generan energía y, por tanto, no es una estrella; es lo que llamamos una enana marrón. La masa más grande que puede tener una estrella no está bien determinada. Sospechamos que existe una masa superior, porque hay procesos físicos que hacen muy difícil la formación de estrellas muy masivas, pero puede que este límite dependa de otros factores, como la composición química de las propias estrellas. Sabemos que existen estrellas con unas 100 masas solares, porque las hemos pesado (es una manera de hablar; determinamos su masa gracias a su movimiento orbital y la ley de la gravedad). Muchos astrónomos creen que en nuestro Universo local la mayor masa que puede tener una estrella es alrededor de 150 masas solares, aunque otros especialistas elevan este número hasta 300 masas solares.

Así que la masa del Sol tampoco es particularmente llamativa. Pero, antes de deprimirse por la vulgaridad de nuestra estrella, cabe preguntarse ¿cuántas estrellas hay de cada tamaño? Ya hemos dado una pista al decir que es muy difícil que se formen estrellas muy masivas. Para saber por qué, tendremos que pararnos un momento a aprender cómo se hace una estrella.

Las estrellas nacen en el interior de grandes nubes de gas y polvo que habitan el medio interestelar [3]. Estas nubes originalmente presentan densidades pequeñísimas y temperaturas muy bajas (unos pocos Kelvin). De repente y por una variedad de motivos que no siempre entendemos bien, las nubes se fragmentan en pequeños trozos que se contraen en respuesta a su propia gravedad hasta tener en su interior presiones y densidades tan altas que se producen reacciones nucleares. La fragmentación está dominada por procesos físicos que operan mejor a ciertas escalas que a otras. Como consecuencia, resulta más sencillo formar estrellas con ciertas masas que con masas mayores o menores.

Si suponemos que una nube dada crea una población coetánea de estrellas - una aproximación un poco burda, porque, entre otras cosas, el tiempo de formación de las estrellas también depende de su masa (en este caso, de manera inversa) –, la distribución en masas de las estrellas será la consecuencia última de todo el proceso. Al cabo de un corto tiempo (astronómicamente hablando) no quedará ningún rastro de la nube que formó las estrellas. La distribución en masas será la única característica observable que nos deje el episodio de formación estelar y, por tanto, la única medida que podamos hacer que nos dé información sobre cómo se ha producido este episodio. La ley que nos dice cuántas estrellas de una masa dada pertenecen a una población se llama función inicial de masas. Su cálculo en diferentes regiones de formación estelar es una de las actividades más importantes para muchos astrofísicos, porque creemos que, sabiendo las

condiciones iniciales y teniendo una descripción del resultado final, podremos entender cómo sucede la formación estelar.

Como consecuencia de estas mediciones, sabemos que hay una masa estelar preferida por la naturaleza. Independientemente de las condiciones iniciales, nunca parece estar muy lejos de un 50% de la masa solar. Si una nube se fragmenta y surge una población estelar, casi todas las estrellas que la compongan tendrán masas bajas y habrá una mayoría con masas entre un 20 y 60% de la del Sol. Para masas más altas, se comprueba que el número de estrellas que se forma es inversamente proporcional a su masa. Pero esta proporción no es lineal, sino exponencial. El número de estrellas disminuye de manera tremendamente rápida conforme la masa crece, siguiendo lo que se conoce como Ley de Salpeter. Así, por ejemplo, por cada estrella de 30 masas solares que llega a nacer, se forman 100 estrellas con la masa del Sol, y muchas más estrellas con masas menores.

Ahora bien, la tasa de formación de las estrellas no es el único efecto a tener en cuenta para saber cuál es la masa típica de las estrellas. Las estrellas mucho mayores que el Sol viven vidas mucho más breves. Volviendo al ejemplo, una estrella de 30 masas solares vive aproximadamente 6 millones de años, pero una estrella de masa solar vive unos 10.000 millones de años. Las estrellas más masivas son escasísimas, aunque resulten moderadamente fáciles de encontrar por su brillo descomunal. Así nuestra pregunta inicial se contesta con una aparente paradoja: aunque la masa del Sol es relativamente baja para una estrella, casi todas las estrellas que existen en la actualidad son menos masivas que el Sol. El Universo está lleno de estrellitas.

17.3 Hacia una teoría de la formación estelar.

La Astrofísica es una ciencia que se ocupa de cuestiones fundamentales. Hoy en día, uno de sus objetivos principales es entender cómo se forman las estrellas. Tenemos teorías bastante avanzadas sobre estructura estelar y una teoría de evolución estelar [1] que reproduce con bastante acierto muchas propiedades de las estrellas en el cielo. Sin embargo, aún no tenemos una teoría completa de la formación estelar y, de rebote, la formación de sistemas planetarios alrededor de las estrellas. Esta deficiencia se debe, en buena parte, a la enorme complejidad del proceso. Para entender bien la formación estelar necesitamos comprender simultáneamente los mecanismos físicos que actúan a la escala de la nube gigante de gas (cientos de años luz) y a la escala de una estrella. Ningún ordenador, por potente que sea, es capaz de calcular variables físicas manteniendo la homogeneidad sobre escalas espaciales que cubren un factor de casi mil millones.

Una segunda dificultad es observacional. Como la formación estelar se produce en el interior de las nubes de polvo, no alcanzamos a verla. Los sitios en los que tiene lugar el nacimiento de las estrellas aparecen como manchas negras en las imágenes del cielo. Por suerte, esta complicación se vuelve menor cada día gracias al desarrollo de nuevos detectores que nos permiten ver en partes del espectro electromagnético donde el polvo es transparente o, por lo menos, mucho menos opaco. Los detectores sensibles al infrarrojo medio a bordo de telescopios espaciales como *Spitzer* o *Herschel* han permitido crear impresionantes mapas de las zonas de formación estelar en nuestra galaxia. En tierra, estamos acabando de construir ALMA, el mayor proyecto de colaboración internacional en Astrofísica, un conjunto de detectores de radiación

submilimétrica que puede escudriñar los detalles del proceso de formación estelar incluso en las regiones más escondidas. Con todo ello y con el siempre creciente poder de computación de los ordenadores, esperamos avanzar en los próximos años hacia una teoría de la formación estelar que tenga poder predictivo.

A día de hoy, podemos explicar de manera cualitativa la formación del sol y otras estrellas y el origen de los cúmulos estelares. Ya existen computaciones que simulan la física de la formación estelar y son capaces de reproducir la función inicial de masas que observamos habitualmente. Sin embargo, aún quedan incógnitas por resolver. Para los observadores, hay dos preguntas muy importantes que contestar. La primera es si vemos siempre una función inicial de masas similar porque en la vecindad solar todas las nubes que forman estrellas son, en realidad, muy parecidas. Quizá si mirásemos en lugares muy diferentes, con condiciones físicas más extremas (como, por ejemplo, en la superficie de contacto entre dos galaxias que colisionan, o en la vecindad de un agujero negro supermasivo) veríamos funciones iniciales de masas distintas. La observación de este tipo de ambientes extremos se lleva mucho tiempo en los telescopios más avanzados en la actualidad. La segunda pregunta (en realidad, muy relacionada con la primera) es si la función inicial de masas ha sido la misma a lo largo de toda la historia del universo. A nivel teórico, se trabaja con denuedo para comprender los numerosos fenómenos físicos que regulan el comportamiento de las nubes a todas las escalas espaciales. Nuestro objetivo es desarrollar una teoría capaz de explicar, partiendo de primeros principios, cómo los componentes elementales se han combinado para formar las estrellas y los seres vivos, es decir, una teoría que nos proporcione una respuesta científica a la ancestral pregunta "¿de dónde venimos?". El estudio del Sol y nuestro sistema solar desempeña un papel fundamental en el desarrollo de nuestras ideas, porque solo aquí somos capaces de apreciar los detalles más finos que siempre nos faltarán, por más potentes que sean nuestros telescopios, cuando escudriñemos los lugares lejanos en que se forman las estrellas. También de este modo, para nosotros, el Sol sigue siendo una estrella especial.

Notas:
1. Ver capítulo 15.
2. Ver capítulo 18.
3. Ver capítulo 11.

Ignacio Negueruela Díez
Doctor en Astrofísica
Profesor Titular en la Universidad de Alicante

Capítulo 18
¿Hasta cuándo podrá mantener el Sol la vida en la Tierra?
(Por José A. Caballero)

La Tierra es un planeta *habitable* y *habitado*. Es habitable porque la vida basada en el carbono se puede desarrollar en el planeta, y es habitado porque de hecho la vida se ha desarrollado en él durante unos 3700 millones de años, aproximadamente un 80% del tiempo desde que la Tierra se formó en el disco protoplanetario del Sistema Solar. Si existe vida “realmente” inteligente en la Tierra, o si un holocausto nuclear o el impacto de un asteroide mayor que el de Chicxulub en el límite K/T afectaría la habitabilidad del planeta son preguntas que no serán discutidas aquí.

Los astrónomos estamos empezando a descubrir planetas habitables alrededor de otros soles. Algunos de estos exoplanetas podrían estar habitados, y para determinarlo estamos desarrollando telescopios e instrumentos que podrían demostrarlo en el futuro. Mientras tanto, los astrónomos usamos el término habitable de una manera bastante laxa. La zona de habitabilidad la definimos como el rango de separaciones a su sol en las que la superficie de un exoplaneta telúrico puede mantener agua líquida durante largos periodos de tiempo. Por planeta telúrico entendemos aquel que se parece a la Tierra: ni demasiado pequeño, para que pueda albergar una atmósfera autosostenida, ni demasiado grande, para que esa atmósfera no se haga tan profunda y densa como las de los planetas helados del Sistema Solar (Urano y Neptuno).

En primer orden de aproximación, a un exoplaneta lo calificamos como habitable si su separación a su estrella (semieje mayor), circularización de su órbita (excentricidad), reflectancia de su atmósfera (albedo) y masa (relacionada con la cantidad justa de atmósfera) se corresponden con la luminosidad de su estrella. En realidad, hay muchos otros factores a tener en cuenta para estudiar su habitabilidad, como el efecto de la actividad estelar (frecuencia e intensidad de fulguraciones estelares), rotación sincrónica del planeta (los más cercanos a su estrella siempre le muestran la misma cara, como la Luna a la Tierra), presencia de campo magnético (que protege la superficie planetaria de la actividad estelar), existencia de tectónica de placas (para que se mantenga el ciclo del carbono), composición de la atmósfera (en Venus hay nubes de ácido sulfúrico), estabilización del eje de rotación (gracias a la Luna, el eje de rotación de la Tierra es relativamente estable)…

La Tierra cumple todos esos requisitos de habitabilidad, y los ha cumplido durante al menos esos 3700 millones de años. ¡Por ahora!

De todos es sabido que el Sol es una estrella enana y que tarde o temprano se convertirá en una gigante roja. En realidad, el futuro a muy largo plazo del Sol (10^{20} años) es convertirse en una enana negra, que es la fase teórica posterior a la de enana blanca. La enana blanca será el núcleo de helio y carbono del Sol futuro, mientras que

las capas exteriores serán expelidas por vientos estelares, formando una preciosa nebulosa planetaria. Antes de esta fase, la Tierra será tragada por la enrarecida atmósfera del Sol cuando este entre en la rama asintótica de las gigantes, dentro de unos 6000 o 7000 millones de años, que es más del tiempo que ha transcurrido desde la formación del Sistema Solar hace unos 4600 millones de años.

¡Pero la Tierra dejará de ser habitable mucho antes! En la actualidad, el Sol quema hidrógeno en su núcleo a un ritmo constante, pero medible. Al transformar hidrógeno en helio, el Sol se va calentando paulatinamente, de forma que el Sol es hoy un 30% más brillante que poco después de que el Sistema Solar llegara a su configuración actual. Dentro de 2800 millones de años, la temperatura de la superficie en los polos alcanzará los 140ºC y los termófilos más resistentes se extinguirán; dentro de 1100 millones de años, los océanos se habrán evaporado, subducido en el manto o fotodisociado por encima de la estratosfera terrestre, y desaparecerán los últimos hongos y plantas fotosintéticas; dentro de 600 millones de años, el 95% de las plantas no podrán realizar la fotosíntesis ya que no habrá suficiente CO_2 disponible en la atmósfera (al aumentar la temperatura, el carbono tiende a fijarse a las rocas carbonáceas de la corteza). Las últimas plantas en morir serán las coníferas; en aquel momento solo quedarán aves, pequeños peces, invertebrados y microorganismos. Los mamíferos seremos los primeros en perecer.

Solo tenemos unos 500 millones de años antes de que nuestra civilización colapse. Tenemos este tiempo para organizarnos y sembrar de vida todos los planetas habitables y deshabitados que encontremos en la Vía Láctea. Como decía el filósofo cordobés Séneca, "*non est ed astra mollis e terris via*" (no hay camino fácil desde la Tierra a las estrellas).

José A. Caballero
Landessternwarte Königstuhl, Heidelberg

Capítulo 19
¿Por qué todos los planetas ocupan aproximadamente el mismo plano orbital?

(Por Luis Felipe Rodríguez Jorge)

Los planos de las órbitas de los ocho planetas de nuestro Sistema Solar y el plano de rotación del Sol coinciden todos dentro de 7 grados. Esta no es una improbable coincidencia sino el resultado de que nuestro Sistema Solar se formó de la contracción gravitacional de una nube de gas y polvo cósmico con rotación. En el plano de la rotación apareció una fuerza centrífuga que dificultó que la nube se contrajera en ese plano. En contraste, en la dirección perpendicular al eje de rotación no se produce esta fuerza centrífuga y la nube, que originalmente era esférica, evolucionó a una estructura aplanada. La masa que se acumuló en el centro de esta estructura daría lugar al Sol. La estructura aplanada y rotante alrededor del protosol se conoce como un disco protoplanetario y de ella se fueron condensando los planetas en escalas de tiempo del orden de unos millones de años.

Hasta recientemente solo teníamos a nuestro Sistema Solar como ejemplo, pero ahora se han logrado imágenes en ondas milimétricas de la emisión del polvo en discos protoplanetarios que nos demuestran que estas aplanadas estructuras son comunes en el Universo y que lentamente se están transformando en planetas.

Luis Felipe Rodríguez Jorge
Doctor en Astronomía
Universidad Nacional Autónoma de Mexico

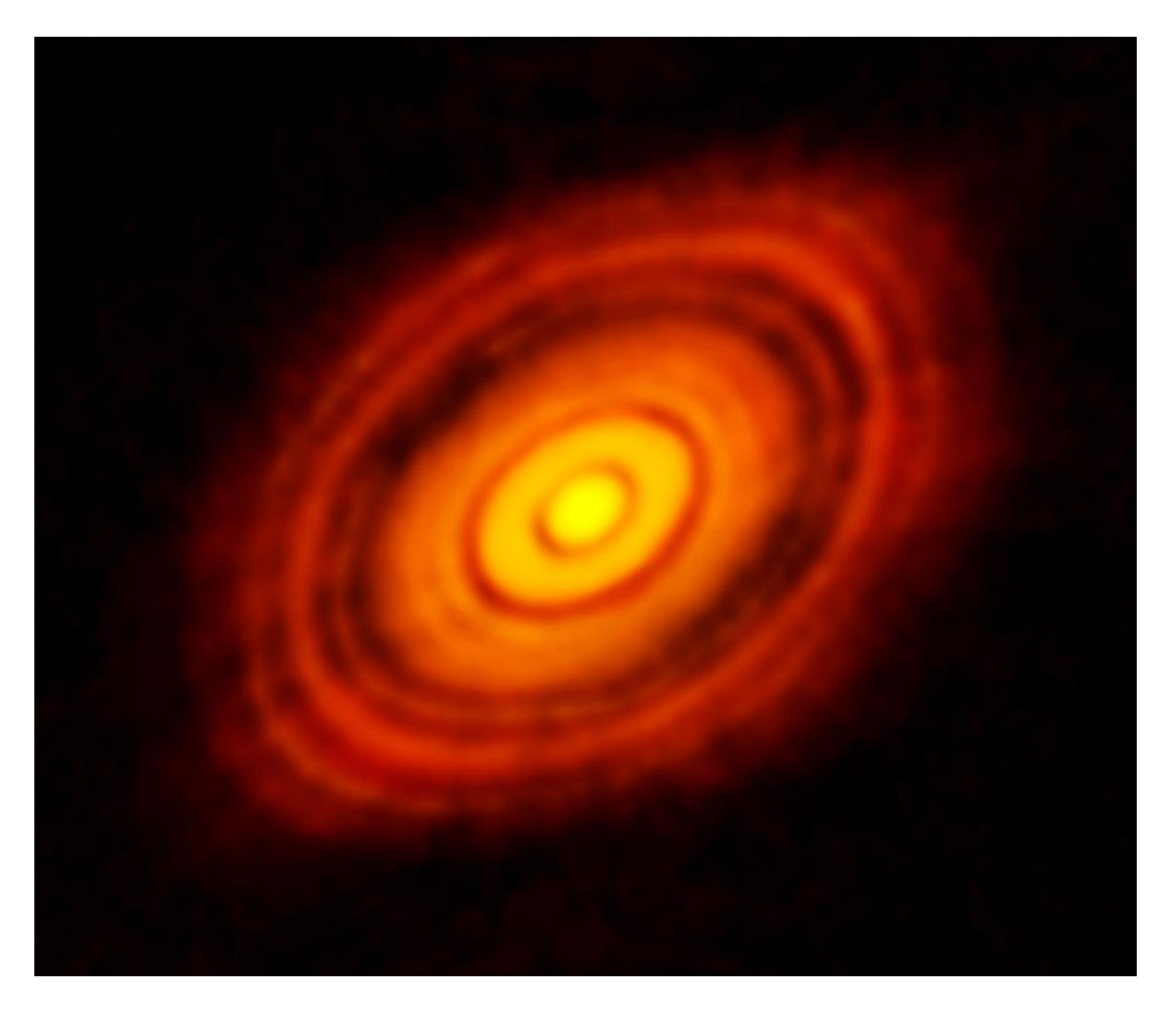


Figura 1: *El disco protoplanetario alrededor de la estrella joven HL Tau. Las observaciones se realizaron a la longitud de onda de 1.3 milímetros con el radiotelescopio ALMA (Atacama Large Millimeter Array). Las brechas circulares podrían indicar la formación de planetas en esas zonas.*

Crédito de la imagen: ALMA (NRAO/ESO/NAOJ); C. Brogan, B. Saxton (NRAO/AUI/NSF).

Capítulo 20
¿Por qué no todos los objetos que orbitan alrededor del Sol son llamados planetas?
(Por Pablo Marcos Arenal)

Si buscamos este capítulo en el libro de Asimov que aquí estamos revisando, nos encontramos con que su título original era "¿En qué difiere Plutón con todos los demás planetas?". Ocurre que, desde la XXVI Asamblea General de la Unión Astronómica Internacional (IAU por sus siglas en inglés), Plutón dejó de considerarse un planeta. Es por esto que era necesario cambiar el título del capítulo y centrarnos en explicar qué es un planeta y en qué se distingue de los demás objetos celestes que orbitan alrededor de nuestro Sol.

Bien podríamos simplemente responder a la pregunta atendiendo a la definición de planeta que nos ofrece el diccionario; Según la Real Academia de la Lengua Española (RAE):
«Planeta: Del lat. planēta, y este del gr. πλανήτης planḗtēs; propiamente 'errante'.
»Cuerpo celeste sin luz propia que gira en una órbita elíptica alrededor de una estrella, en particular los que giran alrededor del Sol: Mercurio, Venus, la Tierra, Marte, Júpiter, Saturno, Urano, Neptuno y Plutón».

Esta definición [1] nos deja en el mismo lugar que nos dejó Asimov allá por 1975 en el capítulo original, con Plutón en la lista de planetas. Conclusión: la definición de planeta es obsoleta y tendremos que atender a la resolución de la IAU para encontrar la respuesta rigurosa que estamos buscando. Lo cierto es que nuestro objetivo aquí no es hacer una disertación lingüística y llevar el tema al terreno de la semántica. Nuestro objetivo es aprender ciencia… y entenderla.

Aun así, todo esto no hace sino plantearnos nuevas cuestiones: ¿Por qué toda esta controversia?, ¿por qué sacar a Plutón de la lista de planetas si sus características no han cambiado y son bien conocidas?, ¿qué ha cambiado para que la IAU tomase esta polémica decisión? Podemos decir que mucha culpa de todo esto la tiene un objeto celeste descubierto por Michael E. Brown y su equipo el 5 de Enero de 2005 desde Monte Palomar. Este objeto, que se bautizó como *Eris*, es más masivo que Plutón (la masa de Eris es de $\sim 1{,}66 \times 10^{22}$ kg frente a los $\sim 1{,}305 \times 10^{22}$ kg de Plutón) y tiene una órbita elíptica alrededor del Sol, por lo que la NASA llego a considerarlo el décimo planeta del sistema solar. Bien podría haberse actualizado la lista de planetas en aquel momento para simplemente añadir a Eris, pero ocurría que poco tiempo antes se había descubierto otro objeto con un tercio de la masa de Plutón y forma elipsoidal que también pudiera haber sido considerado un planeta [2]. La sospecha de que seguirían apareciendo más objetos de este tipo y que la lista de planetas tendría que actualizarse con relativa frecuencia, llevó a la necesidad de encontrar una definición de planeta que fuese inequívoca y aceptada por la mayoría de la comunidad científica.

20.1 ¿Qué es un planeta?

El 24 de agosto de 2006 la IAU acordó en su resolución B5, por amplia mayoría, que un planeta del sistema solar es un cuerpo celeste que:

1) Orbita alrededor del Sol.
2) Tiene suficiente masa para que su gravedad supere las fuerzas del cuerpo rígido, de manera que asuma una forma en equilibrio hidrostático (prácticamente esférica).
3) Ha limpiado la vecindad de su órbita.

Según esta definición, el sistema solar consta de ocho planetas: Mercurio, Venus, Tierra, Marte, Júpiter, Saturno, Urano y Neptuno. Además, se designa una nueva categoría: planeta enano. La única diferencia entre planeta y planeta enano es que este último no ha limpiado la vecindad de su órbita. Por si hubiese alguna duda, la resolución B6 de la IAU zanja: Plutón es un planeta enano.

Así pues, la diferencia entre planeta y planeta enano es que haya limpiado, o no, la vecindad de su órbita. Esto quiere decir que en la vecindad de la órbita del objeto no hay cuerpos de tamaño comparable al suyo, excepto sus satélites u otros objetos bajo la influencia de su campo gravitatorio; se dice en este caso que el objeto tiene *predominancia gravitatoria.* En el caso de Plutón ocurre que está ligado gravitacionalmente con Neptuno, y en su recorrido alrededor del Sol coincide con objetos del Cinturón de Kuiper, como ilustra el dibujo de abajo.

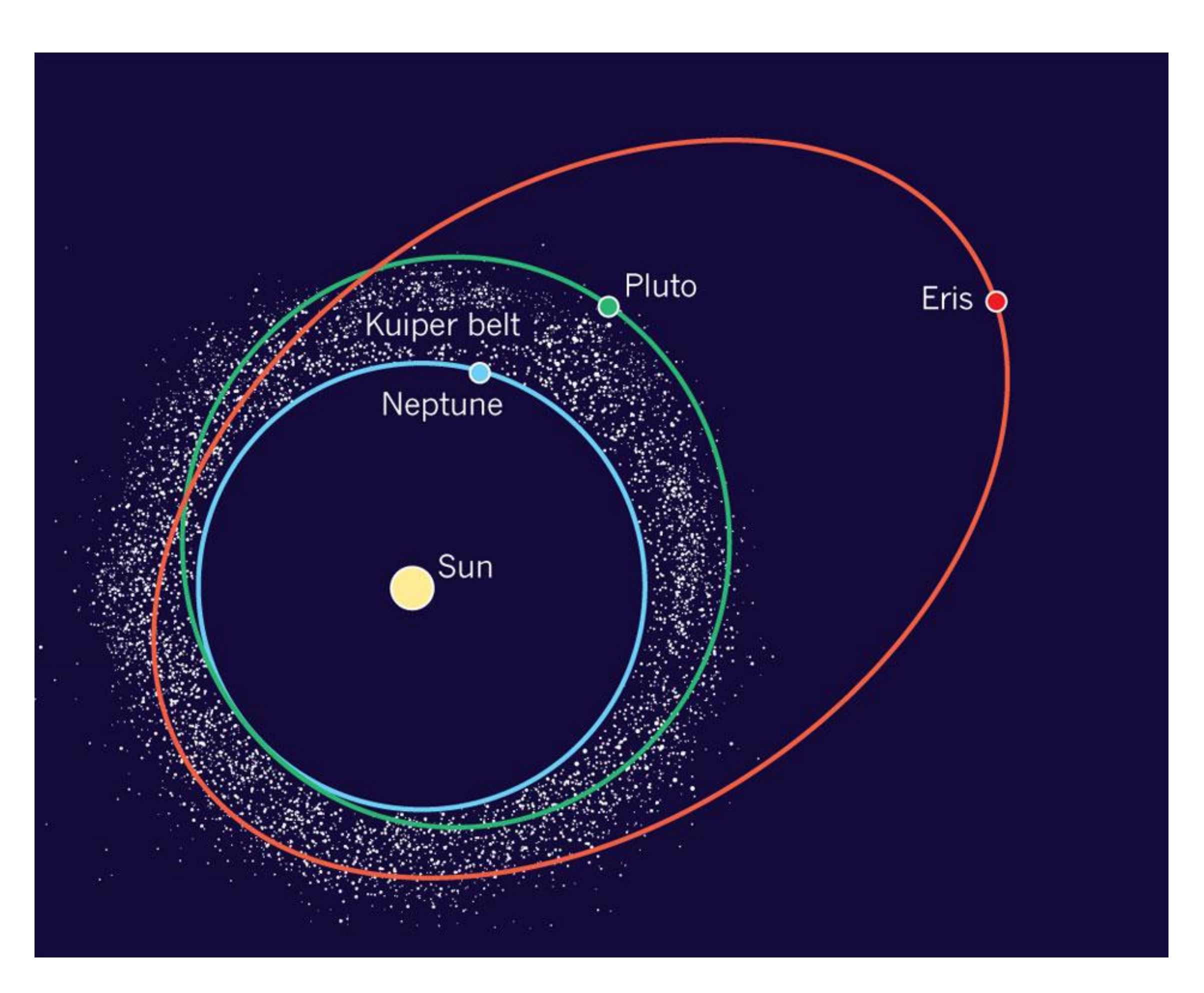


Figura 1: *Plutón, Eris y el Cinturón de Kuiper. Crédito: www.nature.com*

En esta ilustración parece claro que Neptuno debe ser considerado un planeta y Plutón y Eris deben ser considerados planetas enanos, puesto que en la vecindad de sus órbitas hay multitud de objetos del Cinturón de Kuiper. Pero surgen nuevas preguntas: ¿y si fuese Plutón el que tuviese predominancia gravitatoria sobre Neptuno?, ¿cómo se mide dicha predominancia?, ¿por qué los planetas enanos tienen órbitas más excéntricas que los planetas propiamente dichos?, ¿no sería más fácil discriminarlos atendiendo a un criterio de excentricidad?

En realidad no existe un criterio oficial que determine cuando un planeta ha limpiado la vecindad de su órbita o tiene predominancia gravitatoria, aunque algunos astrónomos sí han estudiado la cuestión y ofrecen diferentes criterios para proporcionar unos valores cuantificables que permitan hacer la distinción de manera inequívoca. Los criterios más conocidos son los de Stern–Levison, Soter y Margot, aunque lo cierto es que un planeta nunca puede limpiar completamente su vecindad orbital porque a lo largo de su vida aparecerán asteroides y cometas que interaccionen con el planeta al cruzar su órbita y puedan perturbar su predominancia gravitacional con el paso del tiempo. Atendiendo a cualquiera de los criterios mencionados, utilizando las masas y órbitas de los candidatos a planeta, no existe ninguna duda de que los planetas deben ser los 8 propuestos por la IAU, y que es Neptuno el que tiene predominancia gravitatoria sobre Plutón y no al revés.

Por otro lado, la excentricidad de una órbita no tiene porque estar ligada con la predominancia gravitatoria. La excentricidad de una planeta nos dice lo cerca que está de tener una órbita perfectamente circular (excentricidad igual a cero). La excentricidad de las órbitas de los 8 planetas es menor de 0.1 (oscilan entre 0.007 de Venus y 0.093 de Marte) excepto la de Mercurio, que es de 0.206. Por tanto, al tener Mercurio una excentricidad mayor que la de los planetas enanos Ceres (0.080), Haumea (0.189) y Makemake (0.159), queda claro que no se puede utilizar como criterio para discernir entre unos y otros.

20.2 Equilibrio hidrostático

Habiendo aclarado el tercer punto de la resolución de la IAU, y dado que el punto primero no necesita mucha aclaración, nos queda explicar el segundo punto. En este punto nos encontramos con otro concepto al que debemos prestar atención: ¿qué quiere decir que su gravedad supere las fuerzas del cuerpo rígido y asuma una forma en equilibrio hidrostático? Pues simplemente que tiene forma (aproximadamente) esférica. De nuevo, se puede entrar en la discusión de qué es aproximadamente esférico, pero resulta evidente que ninguno de los planetas puede llegar a ser una esfera perfecta. En un cuerpo de tamaño planetario y en un entorno como el del sistema solar (llegan más de 40.000 toneladas de material a la Tierra cada año) siempre aparecerán perturbaciones que le alejen de ser una esfera matemáticamente perfecta.

Podemos considerar un objeto celeste con una masa cualquiera, pero no con una forma cualquiera; imaginemos el típico satélite artificial con sus paneles solares desplegados a los lados de su cuerpo: tiene la forma que los ingenieros han considerado que es óptima para su función, muy lejos de ser esférica. Consideremos que éste es un objeto celeste, orbitando alrededor del Sol ¿podría ser considerado un planeta? Cumple el punto 1, pero incumple los puntos 2 y 3 que exige la IAU para serlo. Incumple el

tercer punto porque no habría limpiado su vecindad orbital. Como ya hemos visto, no puede tener predominancia gravitatoria pues solo tiene unas pocas toneladas de masa y acabaría viéndose atrapado por la atracción gravitatoria de otro cuerpo más masivo. Pero, ¿qué ocurriría si tuviese la misma forma que hemos propuesto pero una masa muy superior? Imaginemos que pudiésemos construir un satélite mucho más grande y masivo, y que podemos lanzar este inmenso satélite al espacio y colocarlo en la órbita de Mercurio (por ejemplo). Pongamos también que tiene una masa mayor que la de Mercurio, de manera que fuera nuestro satélite el que tuviese predominancia gravitatoria. Ahora sí estarían cumpliéndose los puntos primero y tercero, pero seguiría sin cumplir el segundo por mantener su forma original, lejos de ser esférica, y no podría considerarse un planeta. En realidad este sería el menor de los problemas de nuestro satélite, puesto que un objeto de tal masa con una forma irregular, acabaría colapsando sobre su propio centro de masas, rompiendo las fuerzas de sólido rígido y adoptando un estado de equilibrio hidrostático. Las estructuras que soportan los paneles solares colapsarían por su propia gravedad y caerían como una torre de fichas de dominó que hubiésemos hecho demasiado alta. Y lo mismo ocurriría con cualquiera de las partes que sobresaliesen de su irregular estructura.

Tendría que pasar algún tiempo hasta que nuestro maltrecho satélite llegase a tener forma aproximadamente esférica, pero es lo que acabaría ocurriendo. Un cuerpo celeste con una mayor constitución gaseosa que un cuerpo rocoso tardará mucho menos en adoptar una forma esférica, pero en el tiempo necesario para el proceso de formación planetaria es lo que acaba ocurriendo tanto con uno como con otro.

Existen discrepancias respecto de si la definición de planeta que propuso la IAU y que aquí hemos visto es la más apropiada, y hubo muchas protestas cuando se eliminó a Plutón de la lista de planetas. En cualquier caso, esta definición es actualmente aceptada por la inmensa mayoría de la comunidad científica y permite determinar inequívocamente si un objeto celeste es un planeta o no, por lo que a partir de ahora, si descubres algún nuevo cuerpo celeste en nuestro sistema solar, podrás saber si es o no un planeta. ¡Mucha suerte en la búsqueda!

Notas:

1. Podría discutirse que otros diccionarios en otras lenguas pueden ofrecer definiciones más rigurosas que la que nos propone la RAE, pero no es el caso de los reputados diccionarios ingleses Cambridge Dictionary y Merriam-Webster (invito a los curiosos a buscar las definiciones que ofrecen e, inevitablemente, a quedar atónitos). Consideraremos la definición de la RAE tan buena como la de cualquier otro diccionario.
2. Existe una polémica sobre el equipo descubridor de este objeto entre los observatorios de Sierra Nevada y Monte Palomar; tanto es así que la IAU no reconoce un descubridor de manera oficial. Aunque la IAU reconoce que el lugar del descubrimiento es Sierra Nevada, el nombre del objeto (el nombramiento es un privilegio otorgado al descubridor) es el propuesto por el equipo de Monte Palomar: Haumea.

Bibliografía:

"IAU 2006 General Assembly: Resolutions 5 and 6" , IAU. 24 de agosto de 2006.

"IAU 2006 General Assembly: Result of the IAU Resolution votes" (Press release), IAU . 24 de agosto de 2006.

Pablo Marcos Arenal

Doctor Astronomía

Investigador del Grupo Aegora de la Universidad Complutense de Madrid

Capítulo 21
Plutón, Plutón… quien te ha visto y quien te ve.
(Por René Duffard)

Plutón era ese noveno planeta que a casi todos nos enseñaron que estaba más allá de Neptuno. A partir de 1992 se comenzaron a descubrir más objetos de su tipo, o sea objetos trans-neptunianos (TNOs en sus siglas de inglés) y también planetas orbitando alrededor de otras estrellas. ¿Dónde está el límite? ¿Cuándo un cuerpo es planeta o asteroide? Hubo que sentarse a discutir la definición de planeta. Plutón fue descubierto en 1930 y clasificado como planeta, luego fue re-clasificado como planeta enano en 2006, unos 76 años de reinado. Plutón, no te quejes, dice Ceres, ex-asteroide, que fue descubierto en 1801 y clasificado como planeta y luego en 1850 re-clasificado como asteroide, para luego en 2006 clasificado nuevamente, pero esta vez como planeta enano.

Plutón fue visitado por la nave New Horizon en julio de 2015, un sobrevuelo muy rápido, de unas pocas horas, después de 9 años de viaje. Las imágenes y datos que envió esta nave cambiaron el concepto que teníamos de Plutón y los TNOs. Sabíamos que Plutón tenía una tenue atmósfera, que tenía manchas de albedo (zonas más brillantes que otras en su superficie), teníamos una estimación de su estructura interior. Sabíamos que tenía un gran satélite llamado Caronte, y 4 satélites más pequeños.

¿Cómo sabíamos todo esto? La presencia de una atmosfera se infiere a partir de ocultaciones estelares producidas por Plutón. Se observa como cambia la luz de una estrella cuando Plutón pasa frente a ella. Cuando se ve que la luz disminuye en forma gradual hasta desaparecer y luego la estrella comienza a verse nuevamente en forma gradual, se infiere que es debido a una atmósfera y hasta se puede caracterizar esa tenue atmósfera que tiene Plutón. Un cuerpo sin atmósfera, al pasar frente a una estrella y producir una ocultación, deja una huella característica, donde se puede ver que la luz de la estrella cae bruscamente al ser ocultada por la superficie del objeto. Al tener atmósfera, la luz de la estrella cae gradualmente en una ocultación.

La estructura interna (si tiene un núcleo más denso, un manto y una corteza) se puede conocer al combinar mediciones muy detalladas de su diámetro y de su masa. Si tenemos el diámetro, tenemos el volumen del cuerpo y teniendo también su masa podemos obtener la densidad media del cuerpo, dato fundamental en ciencias planetarias. El diámetro se puede conocer con buen detalle usando ocultaciones estelares o con mediciones térmicas utilizando telescopios espaciales como Herschel por ejemplo y la masa se determina a partir de la interacción gravitatoria con el/los satélites o con una nave espacial si se puede acercar al cuerpo estudiado. Con esos parámetros se determina la densidad media del objeto y de ahí su estructura interna. La densidad media de la Tierra es de 5.51 gr/cm3, la del agua es de 1 gr/cm3, la de Júpiter, 1.33 gr/cm3 y la de Plutón de 1.88 gr/cm3. Concluimos de esto que la Tierra tiene un interior

formado de rocas, Júpiter como es enorme tiene que contener mucho material en estado gaseoso, y Plutón por su tamaño y densidad tiene que tener un interior que es mezcla de rocas y hielos.

Pero lo más sorprendente que encontró la nave New Horizon, fueron los detalles en la superficie de Plutón. Esas manchas más brillantes que se veían en las imágenes tomadas por el telescopio Hubble, eran debido a la presencia de hielo en parte de la superficie. Pero ese hielo no es de agua, es de nitrógeno. ¿Cómo es que puede existir hielo de nitrógeno? Bueno, Plutón tiene una temperatura media de 40 K, eso es 230 grados bajo cero. ¿Y puede haber tanto nitrógeno? Claro que sí, en la Tierra lo hay, pero está en estado gaseoso, el aire está compuesto de 78% de nitrógeno y 21% de oxígeno. O sea que nitrógeno hay, tanto en Plutón, como en la Tierra, como en los otros planetas.
En Plutón existen casquetes polares o una Antártica compuestos de hielo de nitrógeno, donde se pueden ver montañas compuestas de agua sobre un manto de nitrógeno sólido.
Ese enorme glaciar, nombrado Sputnik por el equipo de la nave, es el más grande del sistema solar. Presenta unas formaciones, como células poligonales (ver figura 1) enormes. Esta capa de hielo de nitrógeno de algunos kilómetros de profundidad, se calienta lentamente por el calor interno del planeta enano y sube lentamente a la superficie como una burbuja de agua hirviendo. Desde arriba, estas estructuras se ven como células poligonales de unos 20 kilómetros de extensión. Es curioso que sea el nitrógeno lo que produce estas células, pero resulta que los 40 K de temperaturas es justo la temperatura donde el nitrógeno puede coexistir como sólido, líquido y gaseoso. Similar a lo que pasa con el agua aquí en la Tierra.



Figura 1: *Zonas poligonales de hielo de nitrógeno en la superficie de Plutón. Cada una de estas zonas tiene unos 20 kilómetros de extensión.*

Estas formaciones están en permanente cambio y se mueven, muy despacio, unos pocos centímetros al año, a la velocidad de crecimiento de una uña, por ejemplo. Esto quiere decir que cada 500 mil años, el hielo de nitrógeno en la superficie de Plutón se renueva completamente. Y esta es la causa principal por la que vemos ese hielo tan brillante o nuevo, y que tiene un albedo alto.

En la superficie de esta planicie de nitrógeno, dentro de estas células de convección, se ven como pequeños cráteres, que son producidos por pequeñas burbujas que llegan a la superficie y explotan para luego congelarse dejando en la superficie miles de pequeños cráteres. Es como ver desde arriba una sopa crema en ebullición y cada vez que una burbuja explota, se congela en la superficie dejando un cráter (ver figura 2).

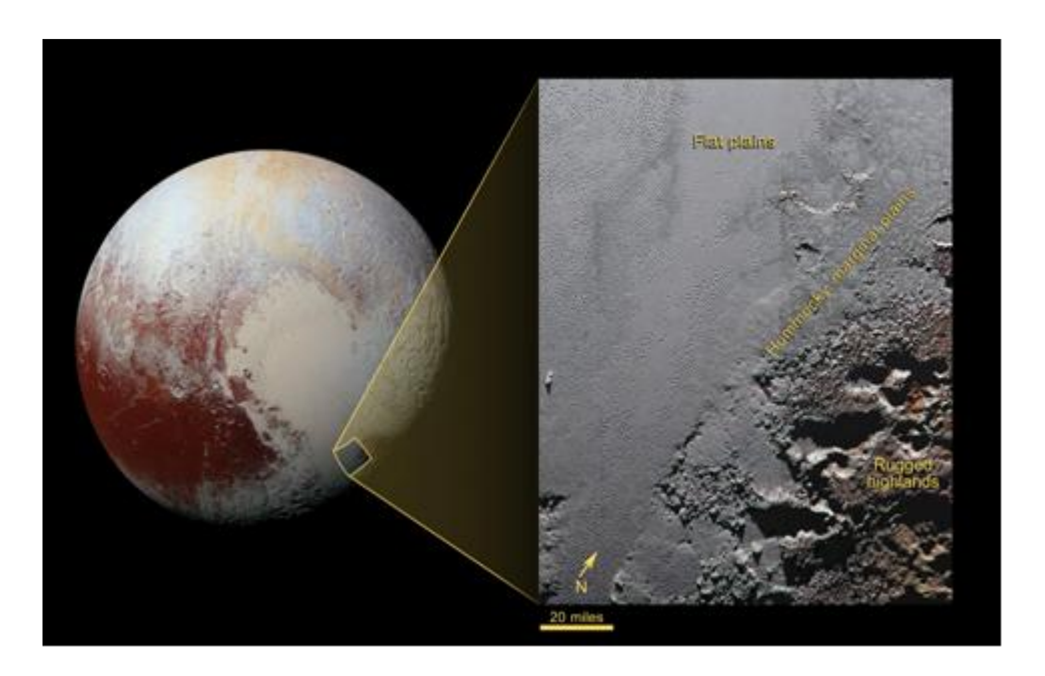


Figura 2: *En esta imagen podemos ver en detalle las zonas planas de hielo de nitrógeno con los pequeños cráteres en el límite con las regiones más oscuras donde acaba el casquete de hielo. Y en este link un video de un sobrevuelo con detalles increíbles sobre la superficie de Plutón*
https://www.nasa.gov/feature/new-horizons-best-close-up-of-plutos-surface

Me imagino un diálogo entre dos geólogos en un congreso de un futuro próximo:

-¿Tu a qué te dedicas?

-A la geología de las rocas volcánicas, como las que hay en Tenerife, Hawai, Marte, la Luna, Mercurio, el asteroide Vesta. ¿Y tú?

-Yo me dedico a la geología de los volátiles. Glaciares de N2, montañas de agua, crío-volcanes, células de convección de N2, planicies de metano congelado.

-¿Tomamos un café caliente?

-No gracias, mejor un granizado.

Y volviendo a la atmosfera de Plutón, ¿es parecida a la atmósfera de la Tierra? Para nada. Primero porque la presión atmosférica es muy, pero muy inferior. Algunos

expertos se niegan a catalogar esto como atmósfera (ver figura 3). La composición es fundamentalmente nitrógeno, con algo de metano y algo menos de monóxido de carbono. Y tiene una presión atmosférica de unos 10 microbares (la presión atmosférica a nivel del mar aquí en el Tierra es de 1 bar). Esta atmósfera existe rodeando a Plutón como una neblina solo porque este planeta es suficientemente grande como para retenerla. Los gases de la atmósfera están ahí debido a la sublimación de los hielos de la superficie.



Figura 3: *Detalle de la superficie de Plutón donde también se aprecian las diferentes capas en la atmosfera.*

¿Y qué podemos decir sobre Caronte? Es el satélite más grande de Plutón, de hecho son un sistema binario. La relación de tamaños entre Caronte y Plutón es mayor que la relación de tamaños entre la Luna y la Tierra.

Se cree que Caronte se formó a partir de una colisión de otro cuerpo con Plutón. Todavía se ven marcas de esa colisión en Caronte que no ha conseguido borrarlas de su superficie (ver la imagen de la figura 4). En esa imagen se puede apreciar una fractura, como el Gran Cañón del Colorado, pero mucho más profundo. La imagen se obtuvo en el mayor acercamiento de New Horizon a Caronte, y se puede apreciar justo en el limbo, la profundidad de la fractura. Tiene unos 9 km de profundidad y unos 700 km (casi toda España de norte a Sur) de largo. Este cañón es casi tan profundo como el Valle Marineris en Marte. Se cree que este cañón se formó al congelarse un océano en el interior de Caronte y se expandió generando esa súper fractura.

Plutón tiene otras 4 pequeñas lunas: Styx, Nix, Kerberos eHydra. Antes de la visita de la nave New Horizon, se conocía muy poco de ellas, gracias al telescopio espacial Hubble. Ahora se conoce algo, poco más, como que tienen solo unos 30 a 40 km de diámetro y que son bastante brillantes, tienen el albedo alto. Eso hace inferir que estarían compuestas de hielo de agua. Estos satélites serian fragmentos de la colisión que formó a Plutón y Caronte.

Fue muy difícil obtener todos estos datos, estas fotografías por la nave New Horizon. Desde mi punto de vista es un logro combinado de la ingeniería, astrofísica, telecomunicaciones, software, y muchas otras áreas. Esas fotografías de alta definición que la sonda envió a la Tierra son un logro científico increíble. La nave viajaba a 50000

km/h, la iluminación solar a esas distancias es mucho menor que en la Tierra, el objeto a fotografiar es oscuro, y estaba a 5900 millones de kilómetros de la Tierra.

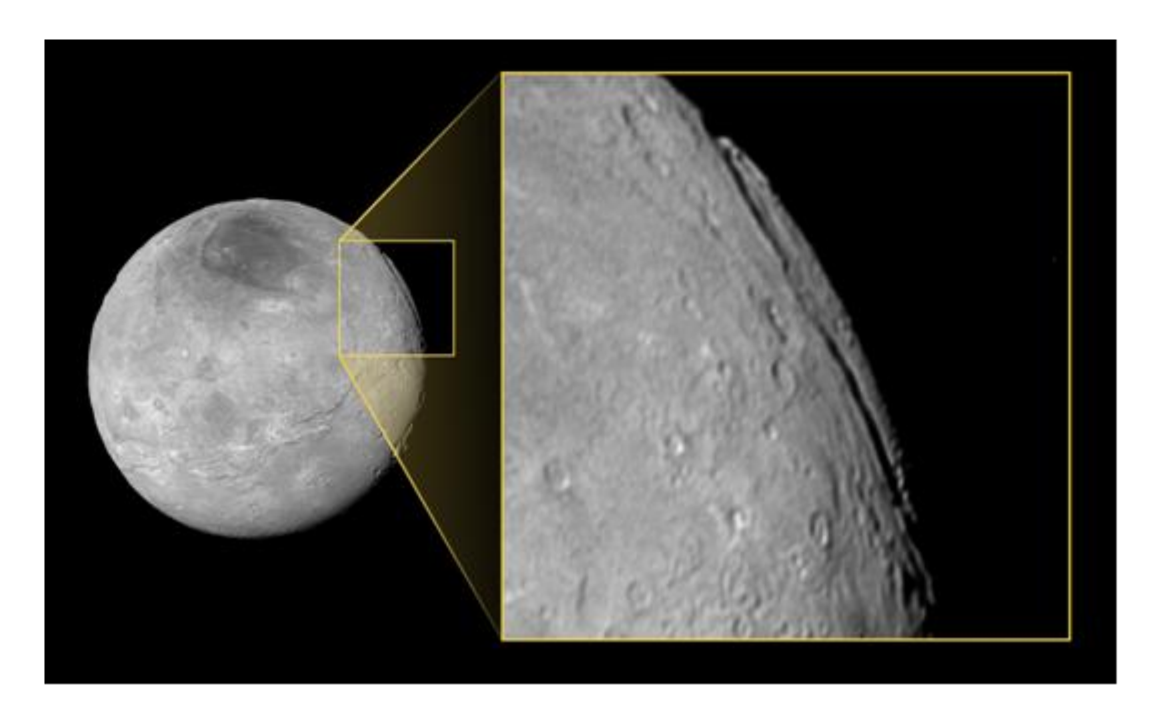


Figura 4: *Caronte visto por New Horizon y en detalle la gran fractura donde se aprecia su profundidad al ser vista en el limbo del satélite.*

¿Con todas esas condiciones quieren que obtenga buenas fotografías? Es como que te pidan construir una cámara de fotos, que luego va a ser sacudida muy violentamente (durante el despegue), que viaje durante 9 años y cada tanto se encienda para obtener fotos de Marte y Júpiter. Luego tienes que enviarle las instrucciones a esa cámara de fotos para programar la serie de fotografías ya que la señal o los datos enviados/recibidos tardan 5 horas en llegar a la nave. Tiene que estar todo programado a la centésima de segundo. Es como estar subido en un coche de Fórmula 1, pasando por un túnel oscuro, y obtener fotografías de un trozo de carbón dentro de ese túnel. Y lo más difícil, programar la adquisición de esas fotografías mientras el coche está viajando hacia ese túnel. Tengo que saber donde está Plutón y sus satélites en aquel momento, como están de iluminados, a que velocidad relativa pasa la nave y a que distancia de la superficie de cada uno de ellos voy a pasar, obtener esos datos, guardarlos en la memoria y luego, cuando todo haya pasado, enviarlos a la Tierra. Sencillamente increíble.

La nave ha enviado 10 Gb de datos desde Plutón (recordemos que una película en formato AVI tiene del orden de 1 Gb de tamaño) y lo hizo a lo largo de 9 meses, es que la velocidad de transferencia es muy lenta. Y luego nos quejamos de nuestra velocidad de bajada de datos en internet.

La nave New Horizon ha marcado un nuevo hito en la exploración espacial. Ha viajado 9 años, para tener un encuentro de unas pocas horas y enviar información sin precedentes de la parte más externa de nuestro sistema solar. Esta visita es la primera a un objeto trans-neptuniano y ha abierto las puertas de una nueva geología. Seguimos

aprendiendo de los datos que ha enviado y se aplicarán a los otros TNOs que estudiemos. Esta parte del sistema solar es la que menos ha cambiado desde su formación y por eso es tan importante estudiarla. Estos estudios permitirán entender también las condiciones que se encuentran algunos exo-planetas que se han descubierto y que se descubrirán. Todo este conocimiento seguro que estará presente en las películas y series de ciencia ficción que se hagan a partir de ahora.

René Duffard
Doctor en Astronomía
Contratado Ramón y Cajal en Instituto de Astrofísica de Andalucía – CSIC

Capítulo 22
¿Por qué siempre vemos la misma cara de la Luna? ¿Pasa lo mismo con los satélites de otros planetas con relación a ellos?

(Por Pablo Marcos Arenal)

En primer lugar vamos a ver qué quiere decir que siempre veamos la misma cara de la Luna desde la Tierra, luego veremos cómo han llegado la Tierra y la Luna a esa situación y por último veremos si esto mismo ocurre con los satélites de otros planetas.

22.1 La cara vista de la Luna

La Luna siempre muestra la misma cara porque su periodo de rotación sobre sí misma es el mismo que el de traslación alrededor de la Tierra, aproximadamente 27.32 días. Al contrario de lo que se creía hasta 1965, no ocurre lo mismo con Mercurio, que tarda 87.9 días en completar una órbita alrededor del Sol pero solo 58.7 días en rotar sobre sí mismo. Por esto desde el Sol sí se puede ver toda la superficie de Mercurio pero desde la Tierra hay una cara de la Luna que permanece siempre oculta. En el caso de la Luna se produce lo que se conoce como acoplamiento gravitacional [1].

La diferencia entre ambos casos es la misma que hay entre una noria y un tiovivo de feria. Si montamos en una noria, un observador que estuviese situado en su eje de giro vería la parte de arriba de nuestra cabeza antes de que la noria comenzase a girar. Cuando la noria hubiese dado un cuarto de giro, este observador nos vería de frente (o de espaldas, dependiendo de la dirección del giro). Y cuando llevásemos media vuelta y estuviésemos en lo más alto de la noria, el observador vería nuestros pies. En tres cuartos de giro vería nuestra espalda (o nos vería de frente). Al completar el giro, de nuevo en la parte de abajo, habría visto toda nuestra “superficie”, de pies a cabeza, literalmente. Habríamos completado una “órbita” alrededor del eje de la noria pero no habríamos girado sobre nosotros mismos. Ahora nos subiremos a un tiovivo mirando, por ejemplo, hacia el centro de la atracción que es donde está el eje de giro y nuestro observador. Cuando hubiésemos completado un giro, este observador solo nos habría visto de frente y no habría podido ver nuestra espalda, al igual que ocurre con la Tierra y la cara oculta de la Luna. En nuestro tiovivo habríamos dado una vuelta alrededor del centro de la atracción (la traslación de la Luna alrededor de la Tierra) y además habríamos dado una vuelta sobre nosotros mismos (rotación de la Luna sobre sí misma).

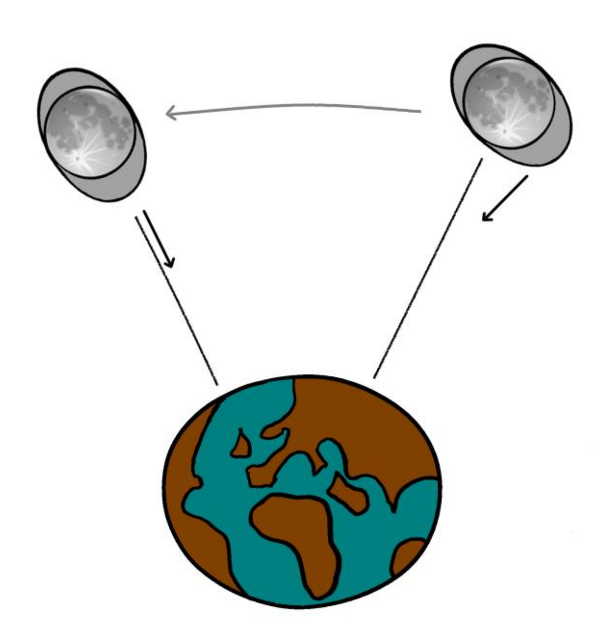
22.2 ¿Siempre la misma cara?

Ahora bien, ¿cómo ha llegado la Luna a sincronizar ambos periodos de giro? En el caso del tiovivo nuestro asiento está fijado al suelo de la atracción forzándonos a dar una vuelta sobre nosotros mismos por cada vuelta que da el tiovivo sobre su eje, pero en el caso de la Luna ¿de dónde sale esa fuerza que la ha llevado a girar sobre sí misma?

Para responder a esto hay que hacer un poco de historia muy muy antigua. Vamos con ello.

La teoría más extensamente aceptada de formación de la Luna asegura que se formó a partir de los restos que se produjeron tras el choque de un cuerpo del tamaño de Marte con la Tierra. De esto hace algo menos de 4.500 millones de años. Tras el choque, la Luna quedó situada mucho más próxima a la Tierra de lo que lo está hoy en día (10 veces más cerca, entre tres y cinco veces el radio de la Tierra ¡imagina el tamaño de la Luna en el cielo!), y los periodos de rotación de Tierra y Luna eran mucho más cortos. En aquel tiempo no había sincronización entre los periodos de rotación de la Luna y su traslación alrededor de la Tierra, por lo que toda la superficie de la Luna era visible desde la Tierra. Además, debido a la proximidad entre los dos cuerpos, la fuerza de la gravedad era muy superior, por lo que la forma de la Luna era más ovalada de lo que es hoy en día. La forma de aquella Luna primitiva era más parecida a la de un huevo que a la casi-esfera que es hoy en día.

Al no tener forma esférica, la fuerza de la gravedad con la que se atraían Tierra y Luna no estaba equilibrada entre la parte achatada y la parte abultada, de manera que la Tierra ejercía un poco más de fuerza para atraer esa masa extra que sobresale en la parte abultada. En menos de 1000 años, esa fuerza extra que atraía a la parte más abultada hacia la Tierra, llevó a que la Luna modificase su periodo de rotación original (que actualmente no se conoce con precisión) y acabase apuntando con esa parte abultada hacia la Tierra. De esta manera, el periodo de rotación de la Luna sobre su eje se ajustó a su periodo orbital alrededor de la Tierra. Este efecto se conoce como acoplamiento de marea o gravitacional.



En el dibujo de arriba se puede apreciar cómo la línea de fuerza de la parte abultada no es paralela al eje que une los centros de masa de los dos cuerpos. En el dibujo está exagerado este efecto y el abultamiento. Tras muchas órbitas, se llegó a la situación de equilibrio en la que el abultamiento se mantiene orientado en dirección a la Tierra, tal como vemos en la Luna de la izquierda en el dibujo.

Con el paso del tiempo y debido a las fuerzas que equilibrio hidrostático, la Luna ha ido perdiendo su forma ovalada para aproximarse más a una esfera, pero la sincronización entre sus periodos de rotación y orbital alrededor de la Tierra se conserva, por lo que en nuestros días vemos siempre la misma cara de la Luna.

22.3 Otros satélites

Este efecto de acoplamiento de marea no es exclusivo de nuestra Luna, y es habitual en otros satélites en nuestro sistema solar, incluso entre estrellas y planetas. Se conocen incluso casos fuera del sistema solar, como el exoplaneta Próxima Centauri b alrededor de su estrella Próxima Centauri.

En el sistema solar existen muchos casos de satélites acoplados a sus planetas: Marte tiene a Fobos y Deimos en acoplamiento gravitacional; Júpiter tampoco puede ver las caras ocultas de sus cuatro satélites galileanos (ni las de otros cuatro de sus satélites menores) y otro tanto ocurre con quince de los satélites de Saturno, cuatro de Urano y con dos de Neptuno. Como vemos, no podemos decir que el caso de la Luna sea un caso especial.

Plutón y su satélite, Caronte, están acoplados el uno al otro; desde Plutón solo se puede ver una de las caras de Caronte, y desde Caronte solo se puede ver una de las caras de Plutón. Al igual que el planeta enano ejerce una fuerza sobre su satélite, éste la ejerce sobre el planeta enano y, como ya hemos visto, si cualquiera de los dos cuerpos tiene una distribución de masa irregular, esta fuerza tendrá una componente que hace que se produzca la sincronización. Esto nos hace pensar que si la Tierra también tiene una distribución de masa irregular, por estar achatada por los polos y existir los efectos de marea, ¿se producirá el acoplamiento de marea de la Tierra con la Luna de manera que la Luna solo será visible desde un hemisferio terrestre? Pues en efecto, así es, pero para eso faltan 50 mil millones de años, que es mucho más que el tiempo de vida de nuestro Sol. La fuerza que ejerce la Luna sobre los abultamientos de la Tierra hacen que nuestra velocidad de rotación disminuya, de manera que los días se alargan 2.3 milisegundos por siglo.

Este efecto de acoplamiento gravitacional se produce cuando un cuerpo orbita alrededor de otro y tiene una distribución irregular de masa que, como hemos visto, es siempre. Entonces, ¿por qué no ocurre con todos los satélites? La respuesta está en el tiempo necesario para que se produzca este efecto. Cuanto mayor sea la fuerza de atracción entre planeta y satélite (mayor masa y menor distancia entre ellos) y mayor sea el abultamiento del satélite, más rápidamente se modificará el periodo de rotación. Dependerá también de la composición física del satélite (en un satélite gaseoso desaparecerá el abultamiento mucho más rápido que en uno rocoso) y de la velocidad de rotación original del satélite, pero la clave es el tiempo necesario para que todos estos factores lleven al acoplamiento gravitacional: en el caso de la Luna hemos visto que se necesitaron 1000 años, pero puede ocurrir que para otro cuerpo se necesite un tiempo mayor que el tiempo de vida del Universo, y eso es mucho tiempo.

Notas:
1. En realidad el caso de Mercurio es particular porque también existe acoplamiento gravitacional pero en una resonancia orbital 3:2 en lugar de 1:1 como es el caso de la Luna.

Bibliografía:
“How close was the Moon to the Earth when it formed?”, Matija Cuk (http://web.archive.org/web/20130912213336/http:/curious.astro.cornell.edu/question.php?number=699).
“Acoplamiento de marea” (https://es.wikipedia.org/wiki/Acoplamiento_de_marea)
“Solar System Dynamics”, C.D. Murray; S.F. Dermott (1999). Cambridge University Press. p. 184.
“Eclipse predictions and Earth´s rotation”, Fred Espenak (http://eclipse.gsfc.nasa.gov/SEhelp/rotation.html).

Pablo Marcos Arenal
Doctor Astronomía
Investigador del Grupo Aegora de la Universidad Complutense de Madrid

Capítulo 23
¿Volvemos a la Luna?
(Por Ana Inés Gómez de Castro)

La aventura lunar comenzó a finales de los años 50 (siglo XX) con la carrera espacial entre la Unión Soviética (URSS) y los Estados Unidos (EEUU). El 12 de septiembre de 1959 llegaba por primera vez una nave terrestre a la superficie lunar; la nave se llamaba Luna 2 (o Lunik 2) y fue puesta en órbita desde el Cosmódromo de Baikonur por la Agencia Espacial Soviética (RosaviaCosmos). El Luna 2 impactó en la región de Palus Putredinis cercana al cráter Arquímedes.

Entre 1966 y 1969 otras siete naves terrestres llegaron a la superficie lunar: dos lanzadas por RosaviaCosmos, el Luna 9 y Luna 13, y cinco misiones por la Agencia Espacial Estadounidense (NASA), Surveyor 1, 3, 5, 6 y 7. Prácticamente todas las naves alunizaron en el hemisferio norte de la cara visible de la Luna (ver figura 1) aunque en regiones con características muy diferentes de la superficie lunar.

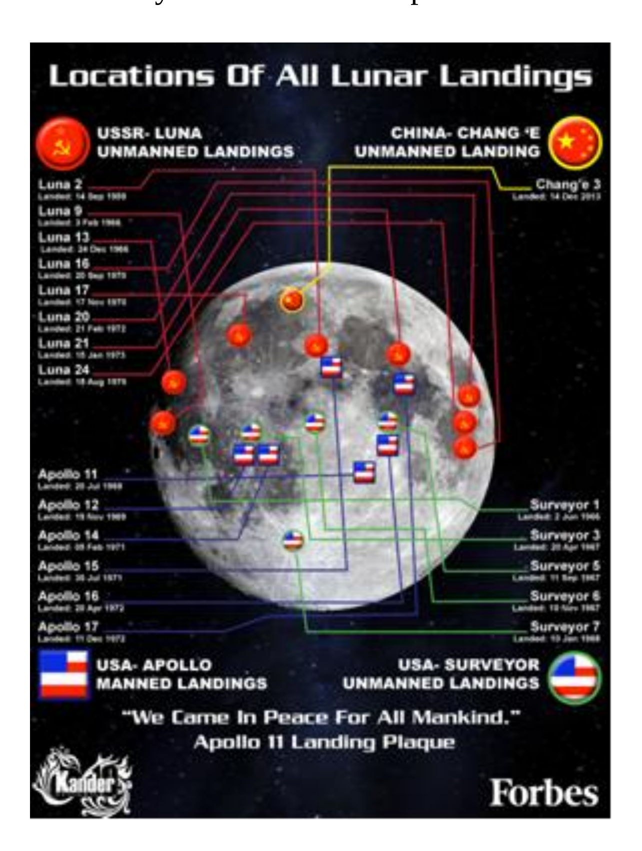


Figura 1: *Zonas de alunizaje de las misiones hasta 2014 [1].*

El gran hito de esta pre-historia lunar tuvo lugar el 20 de julio de 1969, cuando el módulo lunar de la misión Apollo XI (NASA) transportaba por primera vez hasta la

superficie de la Luna a dos seres humanos. La frase de Neil Armstrong: "*un pequeño paso para el hombre pero un salto gigantesco para la humanidad*" (1) forma ya parte del acerbo cultural de la humanidad.

Tristemente, el programa lunar terminó con la carrera espacial. El programa Apollo (NASA) murió en los años 70; la última misión, el Apolo XVII se lanzó en 1972. Por el lado ruso, RosaviaCosmos concluía su programa lunar el 9 de agosto de 1976, con el lanzamiento del Luna 24. A partir de ese momento, el foco de la exploración espacial tomó otro curso: misiones planetarias, observación de la Tierra, desarrollo y operación de la estación espacial internacional, observatorios astronómicos… hasta que recientemente se ha recobrado el interés por la Luna (ver Figura 2).

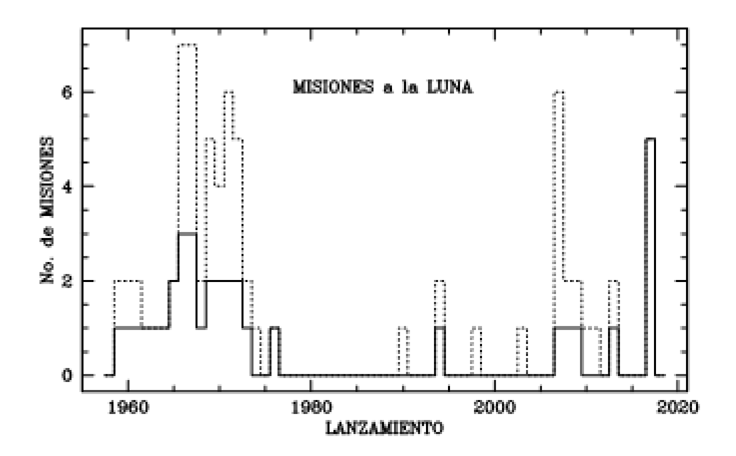


Figura 2: *Misiones lunares anuales desde 1958. La línea continua indica el número de misiones por año que llegaron a la superficie lunar incluyendo misiones que simplemente impactaron con la superficie, robots y los aterrizajes humanos de las misiones Apollo. La línea discontinua marca todas las misiones, incluidos los orbitadores como el Lunar Reconnaissance Orbiter (LRO).*

Los principales responsables de este interés renovado han sido las agencias espaciales más jóvenes: la Agencia Espacial Japonesa (JAXA), realizó una serie de lanzamientos exitosos en 2007, la Agencia Espacial India (ISRO) envió la sonda Chadrayaan-I que descubrió la presencia de agua en la Luna y la Agencia Espacial China (CNSA) inició un robusto programa lunar en 2007 con la misión Chang'e 1, seguida por las Chang'e 2 y Chang'e 3 y que continuará en 2017 y 2018 con las misiones Chang'e 4 y 5.

A las tradicionales agencias espaciales, hay que añadir la actividad de grandes compañías impulsadas por el gigante de internet Google y la visión del grupo Xprize. La competición Google Lunar XPRIZE reúne a equipos que con financiación privada deben poner una nave en la Luna, recorrer con ella 500 metros y transmitir a la Tierra imágenes y video de alta calidad [2]. El primero que lo consiga obtendrá un premio de 20 millones de dólares aunque el objetivo fundamental de los equipos es el desarrollo tecnológico y la búsqueda de una buena posición de partida en la exploración lunar. Los equipos están frecuentemente participados por empresas tecnológicas líderes en sus países de origen y están formados fundamentalmente por equipos científicos y de

ingeniería. En la actualidad hay 16 equipos registrados algunos de ellos con el lanzamiento ya definido y verificado por la organización.

El 27 de septiembre 2003, la Agencia Espacial Europea (ESA) lanzaba la misión SMART-1 inaugurando su actividad lunar. ESA ha mostrado interés en la exploración lunar recientemente, con actividades como retomar los vuelos lunares en estrecha colaboración con RosaviaCosmos y desarrollar la ingeniería para la creación de un asentamiento lunar estable. En los planes conjuntos ESA-RosaviaCosmos está la misión Selene2 que traerá de vuelta a la Tierra una muestra del terreno del polo sur de la Luna.

23.1 ¿Por qué volver a la Luna?

La Luna es rica en recursos minerales y su distribución está siendo determinada por las naves en órbita lunar. Naves como LRO (NASA), Clementine (NASA), Chadrayaan-I (Agencia Espacial de la India, ISRO) han obtenido datos sobre la distribución de rocas hidratadas y minerales en la superficie lunar (ver Figura 3). Las concentraciones más elevadas de metales tales como hierro o titanio, o de algunos elementos utilizables como fuente de energía (p.e. torio que enriquecido puede utilizarse como núcleo fisionable) se encuentran en las cercanías del cráter Copérnico. Otros materiales como el He-3 o el Tritio, son abundantes en el regolito (2) lunar.

Por otro lado, conocer la distribución de rocas hidratadas es fundamental para determinar la localización de futuros asentamientos humanos, así como para la investigación sobre el origen de la vida.

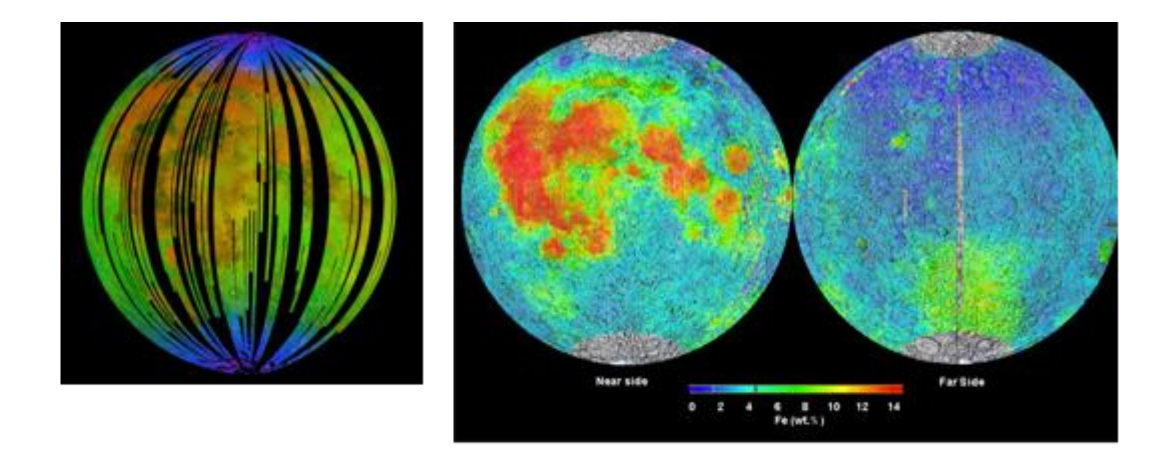


Figura 3: *El panel de la izquierda muestra la distribución de rocas hidratadas en la cara visible Luna; las regiones de colores azules a violeta, situadas en los Polos, tienen una gran concentración de rocas hidratadas [Datos: obtenidos por la misión LRO de NASA]. Los paneles centrales y derechos muestran la distribución de hierro en la Luna, medido por la misión Clementine (NASA) [3].*

En la actualidad, el coste de poner una masa de un kilogramo en la Luna es de aproximadamente 1 millón de euros, haciendo que la explotación minera lunar sea un mal negocio cuando se compara con las explotaciones terrestres, al menos de momento. La explotación mineralógica es una fuerte motivación para la exploración lunar en

especial, a medida que los vuelos se abaraten y los recursos terrestres escaseen. En un futuro más lejano, la abundancia de metales en su superficie, la ausencia de fuentes de oxidación, la disponibilidad de energía y la menor gravedad harán de la Luna un lugar privilegiado como astillero de naves espaciales y como avanzada en la exploración humana del Sistema Solar.

Más allá de las razones económicas existen poderosas razones científicas para instalar laboratorios y observatorios astronómicos en la Luna. La contaminación electromagnética de la Tierra (desde la polución lumínica a las radiofrecuencias) y la ausencia de atmósfera en la Luna hacen de nuestro satélite un lugar privilegiado para la observación astronómica. La Luna es el laboratorio más cercano para el estudio del origen de la vida y la astroquímica; la presencia de rocas hidratadas y el fuerte flujo de radiación ultravioleta proporcionan el laboratorio ideal para reproducir la formación de estructuras complejas y aminoácidos en el espacio.

Pero la razón última para volver a la Luna va más allá de estas consideraciones y es intrínseca a la naturaleza humana. La evolución de nuestra especie se mueve por la curiosidad que nos lleva a la exploración de nuestro entorno. El desarrollo científico y tecnológico está incrementando la esperanza de vida y la población mundial. En consecuencia, tanto la naturaleza humana como la presión demográfica forzarán a los humanos a expandirse en todo el territorio accesible. Y la accesibilidad de la Luna ya está tecnológicamente cercana. En resumen, vamos a volver a la Luna y esta vez para quedarnos.

La exploración y la ocupación de nuestro entorno es también una necesidad para la sobrevivencia nuestra especie: la vida humana es frágil y el Universo hostil. Si nuestra especie quiere sobrevivir tendrá que salir al espacio y extender su área de habitabilidad; es demasiado arriesgado que el único hábitat humano sea el planeta Tierra.

23.2 La ciudad lunar “Nueva Antártida”

La tecnología actual permite enviar con “cierta” facilidad componentes a la Luna pero su coste es todavía muy elevado (1Meuro/kg). Por lo tanto, cualquier propuesta realista pasa por una de estas tres posibilidades:

-Usar como habitáculo la nave de transporte Tierra-Luna
-Desarrollar habitáculos con materiales lunares
-Buscar estructuras geológicas estables y enterradas en la superficie lunar dentro de las cuales se pueda habilitar una base.

La primera propuesta ha sido analizada por varias agencias y solo tiene un inconveniente, el habitáculo debe de ser pequeño, dados los costes de lanzamiento, y habrá de ser “semienterrado” para proteger su interior del viento solar.

La ESA, junto con el equipo de arquitectura Foster & Partners, ha realizado un estudio de la utilización de impresoras 3-D para la creación de edificios en la Luna utilizando como base material lunar. En este trabajo están involucrados estudios y jóvenes empresas españolas del sector de la construcción.

Orbitadores lunares, como el Lunar Recognaissance Orbiter (LRO) de la NASA, están realizando mapas de la superficie lunar con gran resolución y han descubierto zonas protegidas bajo los cráteres y tubos de lava en los que, en principio, podría asentarse una colonia a bajo coste. Aunque el lugar favorito para el establecimiento de una ciudad lunar es el polo sur de la Luna dada la abundancia de agua y las condiciones del terreno (ver Figura 4).

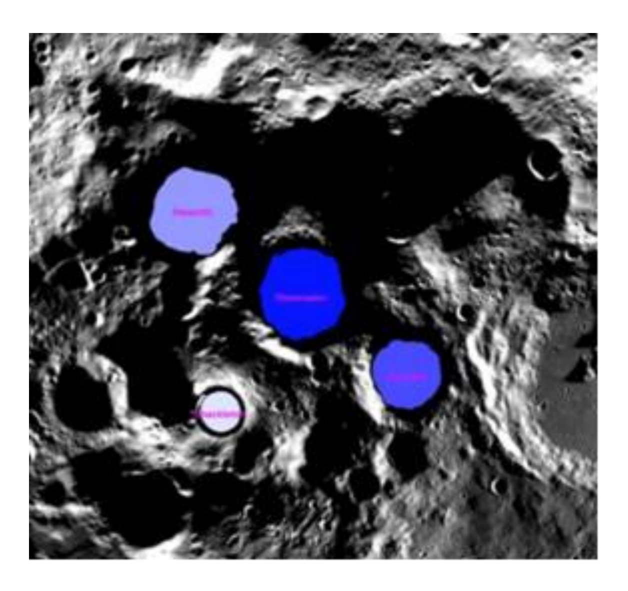


Figura 4: *Imagen del polo sur de la Luna. Los colores en los cráteres marcan la abundancia de hielo del máximo (2%) en blanco a 0.3% en azul oscuro [4].*

Pero, ¿hasta qué punto es acometible el coste de esta aventura? En la actualidad, la comunidad internacional está financiando un gran proyecto espacial: la Estación Espacial International (ISS: International Space Station), ¿serían los costes comparables?

La ISS es el experimento (y la colaboración científico-técnica) más ambiciosa en la historia de la humanidad. La ISS había costado alrededor de 100,000 millones de dólares en 2010, sin contar con los viajes a la estación que incrementarían el monto total a unos 150,000 millones de dólares totales invertidos en el periodo 1998-2010 (3) .

Por lo tanto, y haciendo un prorrateo anual, se podrían colocar unas 10 toneladas al año en la Luna por el mismo costo que la ISS. Pero las Agencias Espaciales están pensando en una aproximación más ingeniosa al problema. Una flota de impresoras 3D que utilicen como material el regolito lunar y como fuente energía el sol, podría crear edificaciones e instrumentación para trabajar remotamente, desde laboratorios terrestres, en la Luna. Obviamente, al final, los humanos nos estableceríamos en la Luna pero ese paso sería mucho más caro y más tardío.

En conclusión, sí, vamos a volver a la Luna y esta vez para quedarnos.



Figura 5: *Una propuesta de ciudad lunar (Moon village) realizada por la ESA, en colaboración con Foster+Partners La ciudad es construida alrededor de módulos enviados desde la Tierra que son recubiertos de regolito Lunar utilizando impresoras 3D [5].*

Notas:
(1) Todavía en 2016, una búsqueda genérica en internet de la frase "es un pequeño paso para el hombre", en español, arroja la frase de Neil Armstrong en las primeras líneas.
(2) El regolito lunar es el material de la superficie de la Luna procesado por el viento y la radiación del Sol.
(3) Esta cantidad es equivalente a un 0,2 % del producto bruto mundial anual (87,25 billones de dólares al año).

Bibliografía:
[1] "The locations of every Lunar landing", Kevin Anderton, July 20th 2015; www.forbes.com
[2] lunar.xprize.org
[3] Lucy et al. 1995, Science, 268, 115
[4] Gladstone et al. 2012.
[5]http://www.esa.int/Our_Activities/Space_Engineering_Technology/Building_a_lunar_base_with_3D_printing

Ana Inés Gómez de Castro
Doctora en Astrofísica
Catedrática de Astronomía y Astrofísica,
Universidad Complutense de Madrid

Capítulo 24
¿Iremos a Marte en un futuro con una misión tripulada?

(Por Jorge Pla-García)

El ser humano es, por definición, un ser vivo que necesita explorar su entorno. Nuestra especie, en su afán por aventurarse en lo desconocido, ha conseguido en los últimos 250 años alcanzar cada medio siglo aproximadamente limites insospechados: entre 1768 y 1779 James Cook exploró el océano Pacifico, entre 1799 y 1803 Alexander Von Humboldt viajó por América Central, entre 1849 y 1855 David Livingston llegó al África austral, en 1911 Roald Admunsen alcanzó el Polo sur, en 1969 Neil Armstrong se convirtió en el primer ser humano en caminar por la Luna y, en la década de 2020 si atendemos a las estimaciones de las empresas espaciales privadas o década de 2030 si se consulta a *NASA*, el siguiente gran hito de la exploración de nuestra especie será pisar por primera vez el planeta Marte. La misión tripulada al planeta rojo será la mayor aventura del siglo XXI y sueños con siglos de antigüedad se verán por fin cumplidos.

"Se buscan hombres para viaje peligroso. Sueldo escaso. Frío extremo. Largos meses de completa oscuridad. Peligro constante. No se asegura el regreso. Honor y reconocimiento en caso de éxito." Este anuncio podría aplicarse perfectamente a la búsqueda de astronautas para la primera misión tripulada a Marte. Sin embargo fue el anuncio publicado en la prensa a principios del siglo XX por Ernest Shackleton para completar su expedición a la Antártida. Se recibieron más de 5.000 solicitudes.

Se suele decir que la primera persona en pisar el planeta rojo ha nacido ya y vive entre nosotros. Me pregunto ¿qué estará haciendo ahora mismo? Probablemente, estudiar. La respuesta a la pregunta de este capítulo, ¿iremos a Marte en un futuro con una misión tripulada?, es rotundamente *sí*. Lo que realmente deberíamos preguntarnos es ¿cuándo? Hoy en día disponemos de la tecnología y los recursos económicos necesarios para realizar una misión tripulada a Marte, una empresa sin precedentes. Sin embargo, se requiere de la voluntad política necesaria para conseguirlo. En este sentido parece que empresas privadas como *SpaceX*, *Boing* o *Blue Origin*, están tomando decisiones mucho más rápido de lo que lo hacen las agencias espaciales públicas. Estas empresas están recuperando el espíritu de la joven *NASA* de los años 60, ágil y osada, y no la conservadora y burócrata de la actualidad.

En el entorno científico marciano se apuesta por una misión privada para poner un ser humano en Marte con una financiación importante de agencias espaciales públicas y el impulso definitivo de una Alianza Internacional. *NASA* ha liderado la carrera durante gran parte de la historia espacial, pero sus limitaciones actuales en cuanto a presupuesto y sobretodo a capacidad de maniobra en cuanto a toma de decisiones independientes al congreso americano es bastante limitada. Y es aquí precisamente donde entran en juego las empresas privadas, las cuales están utilizando los conocimientos adquiridos por las agencias espaciales durante los últimos 50 años, a

la vez que invierten y toman decisiones de una forma mucho más rápida que *NASA*, maniatada por el congreso americano. Se da la paradoja de que parte de la financiación de *SpaceX* proviene precisamente de *NASA*, por lo que todo apunta a que la agencia espacial norteamericana está utilizando la empresa de Elon Musk como plataforma para acelerar los plazos. Tenemos la tecnología, sí. Tenemos los conocimientos, sí. Tenemos la financiación, sí. Solo necesitamos la manera de acelerar los plazos y que no dependa directamente de la aprobación del congreso de EEUU. Por otro lado, y a lo que a financiación se refiere, *NASA* solo puede recibir inversiones públicas, problema del que adolece *SpaceX*, por lo que no sería de extrañar que la futura misión tripulada a Marte estuviera en parte financiada por empresas comerciales como *Coca-Cola*, *Apple* o *Google* por poner un ejemplo. Algunas estimaciones preliminares sitúan en cuatrocientos mil millones de dólares el precio total de una misión tripulada a Marte. Nos encontramos ahora mismo en el momento de la historia espacial donde se esta produciendo un cambio de paradigma y de negocio. El objetivo fundamental de *SpaceX* es que la eterna afirmación "iremos a Marte en 30 años" que se lleva enunciando desde la década de los 60 no se dilate de forma indefinida en el tiempo. Si bien la apuesta de EEUU de poner un ser humano en la Luna fue demoledora, con una inversión para alcanzar el objetivo casi diez veces mayor que la inversión actual (con un máximo de 4.41% del presupuesto federal frente al 0.5% que se invierte actualmente), para el viaje a Marte será necesaria la cooperación internacional. Potencias emergentes en el sector espacial como Europa, China, Japón o India serán imprescindibles para conseguir el objetivo. Me gusta soñar con una Alianza Internacional para la llegada del hombre a Marte.

Para preparar el camino y debido a lo complejo de la empresa, hemos comenzado la exploración enviando sondas robóticas a Marte, que nos proporcionan datos sobre el entorno marciano y sirven como demostradores tecnológicos para la futura exploración humana de Marte.

24.1 Características principales de Marte

Marte, aun teniendo la mitad del tamaño de la Tierra, tiene la misma superficie que esta si tenemos en cuanta de ella solo la superficie firme, sin los océanos. La distancia entre la Tierra y Marte depende de las posiciones relativas de los dos planetas, siendo la distancia media de 225 millones de kilómetros (mínima de 56 millones y máxima de 400 millones de kilómetros). Marte tiene un tercio de la gravedad terrestre, lo cual facilitó la creación de accidentes geográficos enormes, como el Monte Olimpo (tres veces más alto que el Monte Everest) o el cañón Vallis Marineris (10 veces más largo y 6 veces más profundo que el Gran Cañón del Colorado), los más grandes de todo el Sistema Solar, lugares de extraordinaria belleza que podrán ser disfrutados por los astronautas que se desplacen hasta allí. Compuesto por roca y metales, Marte es el planeta del Sistema Solar que presenta una mayor similitud con las condiciones ambientales terrestres, además de rotar aproximadamente al mismo ritmo y en el mismo sentido que la Tierra. Un día marciano, denominado "sol", dura 24 horas y 20 minutos, y su año se prolonga durante 668,59 soles (dos años terrestres), motivo por el cual las estaciones en Marte tienen aproximadamente el doble de duración que las terrestres.

24.2 El clima de Marte: hostil, pero no letal

Marte, que se sepa el único planeta habitado exclusivamente por robots, es un planeta donde corrió agua líquida hace mucho tiempo, pero ahora existe solo en forma de hielo en la superficie y subsuelo y en forma gaseosa en la atmósfera, si bien las efímeras torrenteras, observadas con gran cantidad de sales que podrían hacer descender el punto de congelación del agua, generadas en zonas ecuatoriales durante cortos periodos de tiempo del verano marciano sembraron la confusión, aunque recientemente se ha demostrado que su contenido en agua es bajísimo, apenas del 3%. Desde hace décadas se siguen descubriendo en Marte reservas de agua en forma de hielo, tanto en el subsuelo marciano como en los casquetes polares. Se estima que solo en los casquetes polares de Marte hay tanto hielo de agua acumulado como para generar un océano global de 30 metros de profundidad si se encontrara en fase líquida.

La mayoría de los días en Marte, el tiempo es despejado, soleado y muy frío. En Marte no ha llovido una gota de agua en miles de millones de años. El eje de rotación de Marte es además muy parecido al de la Tierra, con una inclinación de 25º (el de la Tierra es de 23,5º), pero con una excentricidad de su órbita alrededor del Sol mucho más grande (0,093, mientras que en la Tierra es de 0,017) lo cual produce una gran diferencia en la irradiación solar máxima recibida en el hemisferio norte comparada con el sur. Como consecuencia de este diferente aporte energético, las estaciones resultan más extremas en el hemisferio sur. Una muestra de ello es que las temperaturas veraniegas en el hemisferio sur, que raramente superan los 0ºC, pueden ser hasta 30ºC más calientes que las del verano del hemisferio norte.

En cada casquete polar existe una capa de hielo de agua con una capa superficial de hielo de CO_2. Esta última capa se sublima totalmente (pasa de estado sólido a gaseoso) del casquete polar norte en el verano boreal haciendo subir la presión. El proceso inverso se produce en ese momento en el casquete polar sur (invierno austral). El ciclo solar da lugar por tanto a un intercambio de CO_2 entre casquetes. Es como si la atmósfera del planeta respirara con las estaciones.

Existen evidencias que muestran un pasado de Marte con un fuerte campo magnético global, gracias a las observaciones de campos magnéticos residuales locales en el Marte actual llevadas a cabo por orbitadores. El campo magnético de un planeta es su escudo protector frente al virulento viento solar. Cuando Marte perdió su campo magnético se quedó desnudo y a merced del potente viento solar, que barrió su atmósfera, como ha demostrado recientemente la misión Maven, que ha observado como en la actualidad se pierden al espacio más de 8 toneladas de atmósfera al día. Por eso, aunque la atmósfera está compuesta casi en su totalidad de CO_2, la presión media global en superficie es de tan solo ~6 mb (apenas un 1% de la presión terrestre) por lo que la superficie de Marte se asemeja poco a una atmósfera tal y como la conocemos de nuestra experiencia cotidiana. Esta ligerísima atmósfera provoca un debilitamiento del calentamiento por efecto invernadero. Queda patente por tanto que quitarse el casco en Marte sería una autentica temeridad, no solo porque su ligera atmósfera no es respirable para el ser humano, sino porque en esas bajas condiciones de presión el agua comienza a hervir a temperatura ambiente, y las consecuencias para nuestro cuerpo que tiene 65% de agua, serían desastrosas. Para encontrar las condiciones similares de baja presión y baja temperatura de Marte tendríamos que trasladarnos a una altura unas tres veces superior a la de un avión comercial, es decir, hasta la frontera misma del espacio.

Aún así, Marte presenta muchos de los fenómenos meteorológicos que se producen en la Tierra como estaciones asociadas a la oblicuidad planetaria, frentes en superficie, nubes de condensación de vapor de agua y dióxido de carbono, tormentas y remolinos de polvo conocidos como *dust devils* y dunas formadas por actividad eólica, pero sus ritmos de acción son tan lentos que si estuviéramos allí apenas nos percataríamos. Esta tenue atmósfera junto a la ausencia de masas oceánicas, produce una gran amplitud térmica diaria haciendo que la diferencia de temperatura del aire entre el día y la noche sea de hasta 70ºC, por lo que tanto los hábitats como nuestros astronautas tendrán que estar preparados para estos cambios extremos de temperatura. Además, Marte está situado 1,52 veces más lejos del Sol que la Tierra, recibiendo solo un 43% de la luz solar que recibe nuestro planeta.

Existe una incorrecta interpretación de las temperaturas máximas del aire en la superficie de Marte en la literatura y webs de divulgación, en las que la temperatura del suelo a menudo se confunde con la del aire. Si bien el sensor de temperatura de ingeniería (no del aire) en el interior de la cámara frontal *Hazcam* (con recubrimiento metálico) del rover *Spirit* registró una temperatura de +35ºC después de ser estar sometida a una intensa radiación solar, no es riguroso utilizarla como temperatura del aire. Del mismo modo, a menudo se utiliza la temperatura del suelo medida por la estación meteorológica española en Marte *REMS* a bordo del rover *Curiosity* de *NASA* (que en algunas ocasiones puede alcanzar los +15,9ºC) como temperatura del aire. Nada más lejos de la realidad: la temperatura media del aire en Marte cerca de la superficie es alrededor de -55ºC, con valores máximos sobrepasando tímidamente los valores negativos en regiones cercanas al ecuador en los días más calurosos, hasta temperaturas inferiores a -130ºC en los casquetes polares. La temperatura del aire más alta registrada en Marte, en toda la historia de la exploración del planeta rojo, fue la medida por el instrumento *REMS*, que se encuentra el cráter Gale próximo al ecuador en el hemisferio sur marciano, al final de la primavera, momento en el que el cráter recibe mayor irradiación solar. Otras temperaturas máximas del aire cerca del ecuador marciano medidas a la misma altura sobre el suelo son las del aterrizador *Viking 1*, con -17,4ºC y el aterrizador *Pathfinder*, con -11,6ºC. Los rover *Spirit* y *Opportunity*, aunque carecían de instrumentación meteorológica, podían estimar indirectamente la temperatura del aire, obteniendo los registros máximos de -13.15ºC en el caso de *Spirit* y de -8.15ºC en el caso de *Opportunity*. Por lo tanto, aunque la escasez de medidas meteorológicas en Marte hace que estas no sean globalmente representativas, se podría concluir que registrar temperaturas del aire en Marte positivas es algo excepcional.

Marte es un planeta muy polvoriento. El polvo atmosférico tiene un papel protagonista en el clima marciano. Generalmente tiene un efecto anti-invernadero interfiriendo la entrada de energía solar haciendo que disminuyan las temperaturas máximas, pero actúa con un efecto contrario en el infrarrojo actuando como una manta que impide la disipación del calor, haciendo aumentar las temperaturas mínimas. En consecuencia la oscilación térmica diurna disminuye drásticamente. Como curiosidad, las partículas de polvo dispersan la luz solar y producen que los colores en el cielo se inviertan comparados con los de la Tierra: el cielo en Marte es rojizo durante el día y azulado al atardecer. Tormentas locales y regionales de polvo se producen todos los años, generalmente al aproximarse el verano austral. Cada tres años marcianos de media se produce una tormenta global de polvo, que es capaz de cubrir todo el planeta oscureciendo los accidentes más característicos del planeta.



Figura 1: *Autoretrato del rover Curiosity. Sensores de REMS resaltados con círculos negros. Crédito: NASA*

Los peligros para una misión tripulada, lejos de ser los que hemos visto en las películas o en la literatura donde los astronautas son desplazados decenas de metros cuando se encuentran con una de estas tormentas (recordemos que, aunque los vientos pueden alcanzar velocidades de 150km/h, al ser la atmósfera marciana 100 veces más liviana que la terrestre, el astronauta experimentaría una ligera brisa de 15km/h y eso a pesar de que la gravedad marciana es un tercio de la terrestre), está más relacionado con lo finas que son las partículas de polvo, tanto que se podrían comparar con el humo de un cigarrillo. Este polvo, tal y como les sucedió a los astronautas de las misiones Apolo, se introduce por todos los rincones y su peligro no solo esta relacionado con las vías respiratorias y oculares, sino también con la amenaza a los trajes y herramientas que podrían dejar de funcionar, por lo que habrá que prepararse a conciencia para evitarlo.

24.3 Breve historia de la exploración humana de Marte

Durante la historia de la carrera espacial ha habido decenas de planes para la exploración humana de Marte. Como en la carrera espacial a la Luna, rusos y americanos han competido desde los años 60 para alcanzar este objetivo. Problemas con los lanzadores, presupuestos desorbitados, errores de diseño, y el objetivo político cumplido de poner un hombre en la Luna dieron al traste con todos los planes marcianos. Recientemente *NASA* ha vuelto a impulsar la exploración humana de Marte. En Noviembre de 2015 Charles Bolden, administrador de *NASA*, reafirmó el compromiso de la agencia estadounidense para poner un hombre en Marte en la década de 2030 y el esfuerzo para que las siguientes misiones robóticas allanen el camino.

Entre todos los planes destaca uno en concreto en el que se basan la mayoría de los diseños. Además fue el elegido como la arquitectura empleada en la novela, de película homónima, *The Martian*. Esta arquitectura, propuesta por Robert Zubrin y bautizada como *Mars Direct*, pretende abaratar lo que realmente encarece los viajes tripulados a Marte, concretamente la inversión que hay que realizar para hacer el viaje de vuelta, ya que necesitamos combustible para llevar el combustible que nos traerá de vuelta desde Marte. Sin embargo este problema se resuelve en la arquitectura de *Mars Direct* ya que se acometería la misión en diferentes fases: primero se lanzaría a Marte una nave sin tripulación denominada *MAV* (*Mars Ascent Vehicle*), que aterrizaría en Marte y esperaría allí durante un periodo prolongado a los astronautas. El objetivo principal del *MAV* es realizar el viaje de vuelta con los astronautas desde Marte a la Tierra. Durante este lapso de tiempo, el *MAV* se encargaría de forma autómata de fabricar el combustible necesario para el viaje de regreso. Por otro lado se enviarían los módulos de habitabilidad donde vivirían los astronautas. Por último se lanzaría la nave de crucero que transportaría a los astronautas desde la Tierra a la superficie de Marte (viaje de ida). El peso de esta nave sería aproximadamente de 400 toneladas pero como no tenemos ningún lanzador hoy en día capaz de poner esa carga en órbita terrestre (recordemos que el máximo actual son 130 toneladas), se deben realizar varios lanzamientos para ensamblar la nave de crucero en órbita terrestre.

En cuanto a la duración de un viaje tripulado a Marte existen básicamente dos alternativas en función del momento que se elija para el lanzamiento. Si el lanzamiento se produce con Marte en conjunción, es decir, cuando desde la Tierra vemos a Marte en el mismo sentido que el Sol, el viaje es más eficiente porque se consume menos combustible, sin embargo la misión se alarga hasta casi 900 días, con dos viajes de 6 meses de duración y obligando a los astronautas a permanecer en Marte durante 500 días. Si por el contrario el viaje se produce con Marte en oposición, esto es, cuando desde la Tierra lo vemos en sentido opuesto al que vemos al Sol, la misión se acorta hasta los 450 días, con los astronautas viviendo solo unos días o semanas en Marte, a costa de utilizar más cantidad de combustible en el viaje.

24.4 *SpaceX* y la obsesión por colonizar a Marte.

Elon Musk, director ejecutivo y fundador de *SpaceX*, quiere viajar personalmente al planeta rojo (aunque en sus últimas declaraciones ha sembrado la duda afirmando que quiere ver crecer a sus hijos). Para conseguir su obsesivo objetivo necesita abaratar los viajes espaciales. Para ello ha estado estudiando de cerca los diseños de cohetes rusos, ya que son los más sencillos y económicos del mercado. Por este motivo *SpaceX* imprime un carácter de sencillez a todos los diseños, por ejemplo a diferencia de la competencia, solo utiliza queroseno y oxigeno líquido como combustibles. Además, construyen sus propios cohetes, motores y piezas para abaratar costes y tener control absoluto sobre su propia tecnología. Han introducido una novedad que es un hito histórico en la exploración espacial: la reutilización supersónica de la primera etapa del cohete la cual, como si de un tebeo de Tintín se tratara, regresa a su plataforma de lanzamiento después de haber colocado la carga en el espacio. Funcionando a pleno rendimiento y con varios lanzamientos -y aterrizajes- a sus espaldas, ver el regreso de la primera etapa de un *Falcon IX* (capaz de poner 25 toneladas en órbita terrestre) es como estar viendo la grabación del lanzamiento rebobinada. Con esta reutilización los precios descienden considerablemente. Según Musk, el objetivo es abaratar los costes de lanzamiento hasta una centésima parte del

precio final. Esta tecnología de etapas reutilizables sirve también como aterrizaje en Marte.

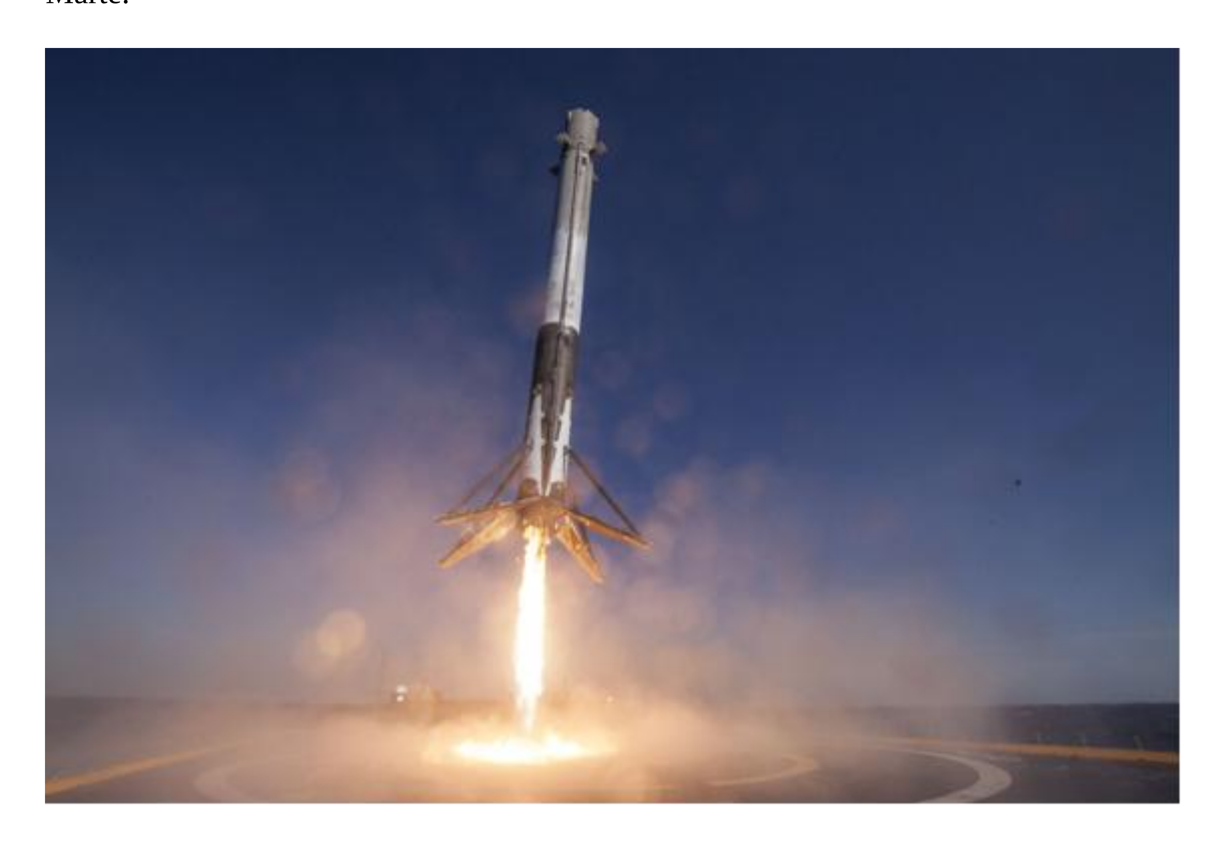


Figura 2: *Imagen del aterrizaje de la primera etapa de un cohete Falcon IX de SpaceX. Créditos: SpaceX*

Elon es un soñador. Ha prometido darle a la humanidad la tecnología para abandonar nuestro hogar y construir uno nuevo.Según él, la colonización de Marte involucrará a cientos de naves reutilizables, cada una llevando entre 100 y 250 colonos, con un precio aproximado de 200.000 dólares por billete. Musk no solo quiere realizar una misión tripulada a Marte, sino colonizar el planeta y convertir a la especie humana en una especie multiplanetaria, para poder tener un hogar alternativo en caso de que una catástrofe amenace nuestro planeta.

El plan consiste en construir el cohete más grande jamás diseñado, con 122 metros de altura y una capacidad para poner en órbita baja 550 toneladas aproximadamente. Utilizará metano y oxígeno líquido como combustible. Tendrá solo 2 etapas: la primera con 42 motores, será reutilizable y pondrá la carga en órbita terrestre y la segunda será la que realice el viaje a Marte, simplificando el diseño. Una vez que la segunda etapa se encuentre en órbita terrestre, la primera etapa realizará entre 3 y 5 lanzamientos (aterrizajes y despegues) para ir cargándola de combustible. La segunda etapa aterrizará en Marte de forma vertical como ya se ha mencionado. La idea es que, a diferencia de la arquitectura *Mars Direct*, sea esta misma nave la que regrese a la Tierra utilizando los recursos locales como combustible.

SpaceX ha desarrollado además, unas cápsulas bautizadas como *Dragon* que, aunque fueron diseñadas desde el comienzo para transportar astronautas, son capaces de enviar mercancías a la *Estación Espacial Internacional* (de hecho es la primera vez en la historia que una empresa privada lo consigue).

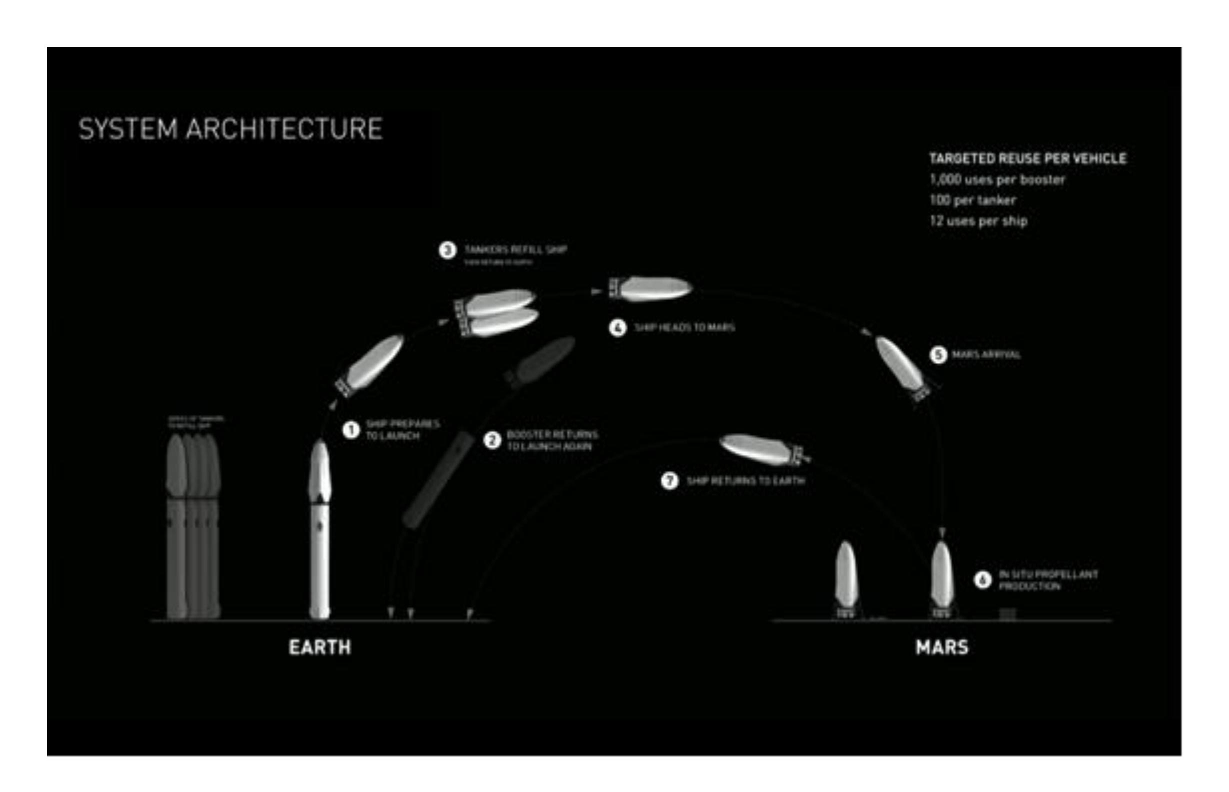


Figura 3: *Arquitectura de SpaceX para la misión tripulada a Marte. Créditos: SpaceX*

En 2018 se utilizará una cápsula *Dragon* modificada y bautizada como *Red Dragon*, para aterrizar en Marte como demostrador tecnológico y sentar las bases de los aterrizajes con cargas pesadas en el planeta rojo, sirviendo como primera fase para la exploración humana de Marte.



Figura 4*: Test de retropropulsores de sonda Dragon de SpaceX. Créditos: SpaceX*

Esta mejora de poder aterrizar en Marte con cargas unas 20 veces más pesadas que las actuales será un salto cualitativo también para las misiones de *NASA* y otras agencias que podrán verse beneficiadas de esta capacidad para incluir más experimentos e instrumentación en las misiones. Esta cápsula *Red Dragon* será lanzada en el futuro cohete *Falcon Heavy* (un *Falcon IX* modificado).

En los planes de *SpaceX* no se mencionan de momento las características de los trajes espaciales que se usarán en Marte, pero estos deberán ser lo suficientemente cómodos como para poder recoger muestras, por lo que no pueden ser muy voluminosos y además deberán estar presurizados debido a la baja presión atmosférica marciana y refrigerados para poder soportar las bajísimas temperaturas. El sistema de soporte vital debe tener un tamaño suficiente para llevar a cabo salidas extravehiculares de larga duración, por lo que deben incorporar generadores de oxígeno. Estos trajes podrían compartir la mayor parte de los elementos de un traje utilizado hoy en día para realizar las actividades extravehiculares en la Estación Espacial Internacional, pero tendrán que tener más movilidad e incluir un sistema de soporte vital compatible con la atmósfera marciana.



Figura 5: *Recreación artística del cohete Falcon Heavy. Créditos: SpaceX*

24.5 Utilización de recursos in situ

La utilización de recursos in situ (*ISRU* en sus siglas en ingles) será fundamental y obligatoria en una misión tripulada a Marte. Debido a que los lanzamientos espaciales se encarecen de forma exponencial cuanto más peso llevan, en todos los planes se asume que las naves tripuladas deben ser lo más ligeras posible y que los astronautas deberán hacer uso de los recursos locales en Marte para autoabastecerse. Los cultivos serán necesarios en el planeta rojo, pero los astronautas se encontrarán con muchos problemas. Al clima hostil y la menor energía solar disponible se le une la mencionada gran escasez de agua en la superficie de las zonas ecuatoriales. Varias son las soluciones propuestas para solventar este problema como por ejemplo purificar la orina de los astronautas para que sea reutilizable, generar agua a partir del oxígeno atmosférico u obtener agua de las reservas de hielo del subsuelo y los casquetes polares.

A estos problemas se suma el poco fértil suelo (regolito) marciano, rico en sales llamadas percloratos que en teoría dificultarían el cultivo de cualquier vegetal terrestre. Sin embargo, algunos estudios previos sugieren que se pueden minimizar los efectos de los percloratos fácilmente empapando previamente el suelo en agua.

Los habitantes que viviesen en Marte tendrían que producir metano y oxígeno con el agua y los depósitos de carbón que se han encontrado bajo la superficie del planeta. El oxígeno será utilizado no solo para respirar, sino también como combustible para el viaje de vuelta.

NASA esta invirtiendo muchos esfuerzos en proyectos *ISRU* y como muestra de ello es la elección de los instrumentos del nuevo rover marciano *Mars2020* (en el momento de la publicación de este libro todavía no ha recibido nombre) que se lanzará en el año 2020 y que servirá como demostrador tecnológico para soporte vital y combustible de vuelta para abrir el camino de la exploración humana de Marte. Esa es una de las tareas del instrumento *MOXIE*, cuya misión es fabricar oxígeno no solo para el consumo humano, sino también para hacer volar en un futuro a la nave que llevará a los astronautas de vuelta a la Tierra. El oxigenador es muy importante porque la atmósfera marciana está compuesta casi exclusivamente por dióxido de carbono, y los motores no funcionarían sin la combustión que el oxígeno hace posible. Otro ejemplo *ISRU* es la ambiciosa estación meteorológica española *MEDA*, que irá también a bordo del rover *Mars2020*, cuyo objetivo será la caracterización de las condiciones medioambientales para que los planificadores de misiones puedan proteger a los futuros exploradores humanos.

24.6 Peligros para la salud del astronauta

Ya hemos hablado de lo hostil que puede llegar a ser el planeta Marte. Frío, desolado y escaso en recursos vitales. Muchos argumentan que nuestra especie no esta preparada para sobrevivir fuera de la Tierra. Es importante recordar que no hace mucho tiempo las mejores mentes del mundo pensaban que un viaje a la Luna era totalmente imposible.

La gravedad en la superficie de Marte es tan solo la tercera parte de la Tierra. Se desconocen los efectos que puede tener esta baja gravedad en el cuerpo humano, pero sabemos gracias a los estudios realizados en la Estación Espacial Internacional, que en ausencia de gravedad los fluidos se desplazan hacia la cabeza presionando el nervio óptico y el globo ocular pudiendo dificultar la visión, además al no tener que luchar contra la gravedad, se debilitan los músculos, entre ellos los del corazón, cuádriceps, espalda y cuello. La piel pierde elasticidad y se vuelve más sensible y las heridas tardan más en curarse, por lo que las infecciones pueden ser más comunes. Los huesos se debilitan considerablemente al perder minerales y calcio, los cuales podrían acumularse y formar piedras en el riñón. Además, la ingravidez confunde al oído, causando mareos y desorientación. Para despejar las incógnitas se están realizando experimentos en la Estación Espacial Internacional, estudiando si nuestro organismo es capaz de sobrevivir a periodos prolongados en el espacio. Recientemente se ha desarrollado un experimento, con la mente puesta en Marte, comparando la evolución durante un año de dos astronautas gemelos, uno de ellos residiendo en la Estación Espacial Internacional y otro de ellos en Tierra para que, pasado el año, pudieran ser comparados y determinar cuales son las consecuencias en nuestro organismo de una estancia prolongada en el espacio. Pasado el año se ha evidenciado que las estancias prolongadas en ingravidez redundan en atrofia muscular y pérdida de masa ósea (aproximadamente un uno por ciento al mes) pero podría paliarse con ejercicio físico o un generador de gravedad artificial en la nave a Marte.

Todos los científicos parecen estar de acuerdo que la mayor amenaza para una misión tripulada es la radiación solar, las tormentas solares y los rayos cósmicos. Esta gran aventura está plagada de peligros, entre ellos los que supone estar sometido a las altísimas dosis de radiación de partículas provenientes del espacio durante el trayecto. Recientes estudios demuestran que la exposición acumulada, para un astronauta con billete de ida y vuelta, sin contar la estancia -más o menos prolongada- en el planeta rojo, equivale a hacerse 33.000 radiografías. Se han propuesto diferentes alternativas para proteger a los astronautas de estas letales radiaciones, por ejemplo construir los hábitats en el interior de cuevas o en el subsuelo marciano que apantallaría de la radiación, utilizar las zonas con campos magnéticos residuales, esconderse de las tormentas solares en las partes blindadas de la nave espacial o módulo de aterrizaje o incluso modificar el ADN, ya que en un reciente estudio se sugiere el empleo de proteínas denominadas "Dsup" que en organismos tardígrados parecen ser las responsables de la gran resiliencia frente a la radiación evitando que el ADN se rompa a medida que la radiación le impacta.

Otra investigación está centrándose en los efectos de la radiación concretamente en el cerebro humano y sus conclusiones elevan aún más la alarma, ya que causan importantes daños cerebrales a largo plazo en roedores, lo que resulta en alteraciones cognitivas y demencia. Los astronautas que viajen a Marte podrían sufrir demencia crónica. Quizás, a su regreso, ni se acordarían del viaje. Investigar cómo afecta la radiación espacial a los astronautas y las maneras de mitigar esos efectos es crítico para los planes de futuras misiones a Marte. Se está trabajando en estrategias farmacológicas que implican compuestos que eliminan los radicales libres y protegen la neurotransmisión de las letales radiaciones.

Otro tipo de inconvenientes, no por ello menos importantes, son los relacionados con problemas psicológicos como pueden ser delirios, demencia senil, depresión, ansiedad o claustrofobia. Para ello se están desarrollando proyectos de simulación de hábitats marcianos en Tierra.

Me gusta soñar con una llegada del hombre a Marte en la década de 2020, en una nave de *SpaceX* financiada por empresas comerciales privadas y agencias espaciales de una Alianza Internacional utilizando la arquitectura de *Mars Direct*.

Jorge Pla-García
Doctorando en Astrofísica (UCM).
Master en Meteorología (UCM).
Master en Astrofísica (VIU).
Master en Ciencia Espacial (UAH).
Investigador del Centro de Astrobiología (CSIC-INTA), Madrid.
Investigador del Southwest Research Institute (SwRI),
Investigador del Scientist, Space Science Institute (SSI),
Boulder (Colorado, EEUU).

Capítulo 25
¿Qué son los ambientes análogos marcianos en la Tierra? ¿Por qué son útiles para buscar vida en Marte?

(Por Alberto González Fairén)

25.1 ¿Qué es un "análogo marciano"?

Décadas de exploración científica han revelado que algunos lugares de la Tierra pueden ser similares a algunos de los entornos geológicos que han sido inferidos (en el pasado) u observados (en el presente) en Marte. Por lo tanto, los análogos terrestres de Marte son lugares de la Tierra que se caracterizan por presentar unas condiciones climáticas, mineralógicas, geomorfológicas y/o geoquímicas similares a las observadas en Marte, en el presente o en el pasado. Estos ambientes análogos terrestres representan una ventana a través de la cual podemos estudiar Marte sin salir de nuestro planeta.

Debemos tener siempre muy presente que cada uno de los análogos de Marte ha de ser comparado con un tiempo geológico marciano preciso y con un lugar específico de la geografía marciana, y que no existe ningún lugar en la Tierra que constituya un análogo perfecto y completo para toda la historia geológica y para toda la superficie de Marte. En cualquier caso, si bien cada uno de los análogos marcianos por sí solo no puede representar adecuadamente los cambios en las condiciones ambientales de Marte a través del tiempo, sí que podemos extraer lecciones valiosas sobre la posible habitabilidad de Marte en el presente y en el pasado mediante la combinación de la investigación de estos análogos con simulaciones realizadas en laboratorio y con los datos obtenidos a través de los orbitadores y las sondas que han aterrizado en Marte.

El estudio de análogos marcianos vinculados a la historia geológica de Marte nos permite comprender mejor los procesos que han dado forma a la superficie del planeta, incluyendo su evolución geológica y geoquímica. También nos facilita la evaluación de la diversidad de ambientes en la superficie y en el subsuelo, tanto en el espacio y como en el tiempo; nos ayuda a determinar la mineralogía y el origen de ambientes únicos, que son importantes para la extrapolación de los resultados a Marte; y nos permite evaluar el potencial para la aparición de la vida y la preservación de los diferentes tipos de biomarcadores fósiles a través de la historia del planeta.

Muchos análogos terrestres reúnen condiciones geoquímicas que los hacen habitables únicamente para ciertos tipos de organismos muy especializados y adaptados a estos entornos.

Al intentar comprender la historia de Marte a través del estudio de análogos terrestres, también obtenemos información relevante acerca de los límites de la vida en la Tierra, así como de la capacidad de los organismos para adaptarse y sobrevivir en ambientes donde el agua líquida se encuentra al límite de su estabilidad.

25.2 Las edades de Marte

La historia geológica de Marte está definida por tres etapas muy diferentes: el Noeico (desde la formación del planeta hasta hace 3.600 millones de años) fue una época que se caracterizó por la presencia de una atmósfera de cierta entidad y abundante agua líquida en la superficie, que formó ríos, lagos, deltas e incluso un océano, en un entorno climático en general muy frío, similar al Océano Glacial Ártico de la Tierra; el Hespérico (de 3.600 a ~ 3.000 millones de años atrás), un periodo de transición aún más frío, y semiárido, durante el que la atmósfera se volvió mucho más tenue, el agua de la superficie estaba en general congelada y los campos glaciares eran habituales; y, finalmente, el Amazónico (desde hace ~ 3.000 millones de años hasta la actualidad), que se caracteriza por ser un período de extrema sequedad y frío intenso, con una atmósfera cada vez más fina, y que ha convertido la superficie de Marte en el desierto árido y extremadamente frío que hoy conocemos.

Estas tres épocas climáticas globales definen también tres posibilidades diferentes desde el punto de vista de la habitabilidad de Marte: una etapa temprana, cuando los requisitos básicos para la vida tal y como la conocemos en la Tierra estaban presentes en Marte (agua líquida y fuentes de energía); una etapa intermedia, durante la que las soluciones líquidas se volvieron escasas y potencialmente no adecuadas para cualquier forma de vida; y la etapa más reciente, durante la cual las condiciones en la superficie han sido en gran parte prohibitivas para la vida, excepto tal vez en algunos nichos aislados.

25.3 Análogos marcianos en la Tierra

Durante las últimas décadas, se han propuesto un gran número de lugares de la Tierra como análogos marcianos, hasta el punto de que hoy se pueden contar por docenas. No obstante, la mayoría presenta un grado de analogía reducido, que se limita a una característica exclusivamente, ya sea climática, geoquímica o mineralógica. Tan solo un pequeño número de análogos marcianos pueden ser considerados realmente como tales, al mostrar un conjunto de rasgos que realmente los asemejan a ciertos paisajes marcianos, presentes o pasados. Entre ellos, los más destacables son:

25.3.1. Análogo del Noeico (I)

La región de North Pole Dome, en Pilbara, Australia Occidental. El cratón de Pilbara se extiende sobre 60.000 km^2, e incluye rocas relativamente poco deformadas que datan del periodo Arcaico (hace entre 3.500 y 2.700 millones de años). El North Pole Dome se sitúa en el centro del cratón, cubriendo un área de 600 km^2, y es una región de abundante vulcanismo antiguo que representa un análogo marciano muy válido, ya que Marte presenta muchas regiones ricas en basaltos poco alterados. Las rocas máficas de Warrawoona incluyen algunos de los terrenos volcánicos más antiguos que se preservan en la Tierra, con una edad comparable a los basaltos de las tierras altas marcianas. Entre los materiales identificados han aparecido filosilicatos similares a las arcillas marcianas, en ambos casos formados por alteración hidrotermal de los flujos de lava basálticos. Algunas de las rocas del North Pole Dome contienen los restos de la biosfera de la Tierra más antiguos que se preservan, en forma de estromatolitos y posibles microfósiles, algunos de ellos con edades superiores a los 3.000 millones de años. El acceso a estos materiales para su estudio está facilitado por la diversidad de los afloramientos y la escasa vegetación de la zona.

25.3.2. Análogo del Noeico (II)

RíoTinto, España. El Río Tinto se encuentra en la Faja Pirítica Ibérica (IPB), una formación geológica en el suroeste de España originada por actividad hidrotermal. El río Tinto fluye a través de los depósitos minerales de la IPB (principalmente hierro y sulfuros de cobre) hasta llegar al Océano Atlántico, dando lugar a la región ácida más grande del mundo. Las aguas del río Tinto tienen una alta concentración de ion férrico que mantiene la acidez de la solución (pH=2,3 de media) a lo largo de toda su longitud, a pesar de los cambios climáticos estacionales y la dilución debida a sus afluentes y la lluvia. La acidez extrema y la alta concentración de metales pesados que se encuentran en el ecosistema del Tinto son la consecuencia directa de los procesos biológicos de los microorganismos oxidantes del hierro y del azufre que crecen en los depósitos de sulfuro metálico de la IPB. Se ha sugerido que las características geológicas del río y su entorno podrían ser similares a las de la región de Meridiani Planum (la zona de exploración del rover Opportunity) durante el Noeico.

25.3.3. Análogo del Hespérico

La isla Axel Heiberg, en el archipiélago Ártico Canadiense. En las regiones del extremo norte del archipiélago Ártico Canadiense existen extensas acumulaciones de hielo masivo. Allí, el permafrost (mezcla de suelo y hielo) es denso y continuo, análogo al permafrost marciano. Estas características están presentes, por ejemplo, en las islas Ellesmere y Axel Heiberg. En Axel Heiberg existen diferentes manantiales de agua líquida que permanecen activos todo el año, incluso durante el invierno, soportando temperaturas muy por debajo de cero grados centígrados; algunos de estos manantiales se mantienen activos por la presión con la que el agua circula bajo la superficie, mientras que otros son de naturaleza hidrotermal. Ambos procesos, y algunos otros como la presencia de salmueras, podrían ayudar a explicar la presencia local de agua líquida en Marte durante el Hespérico. La isla Axel Heiberg, además, proporciona un entorno adecuado en términos climáticos y de presencia de hielo para ensayar los procedimientos de perforación y manipulación de muestras que han de completar nuestros robots en Marte (Figura 2).

25.3.4. Análogo del Amazónico (I)

Desierto de Atacama, Chile. Las evidencias geológicas y la mineralogía del suelo sugieren que las condiciones hiperáridas han persistido en el desierto de Atacama durante al menos 10-15 millones de años; además, el registro sedimentario indica que la región ha sufrido un clima árido desde el Triásico tardío, lo que le convierte en el desierto más antiguo de la Tierra. Hay cuatro características fundamentales que convierten a Atacama en un interesante análogo marciano: 1) los suelos son muy antiguos (hasta 2 millones de años), extremadamente secos, y ricos en un conjunto de sales solubles similares a las encontradas en Marte; 2) muestra niveles extremadamente bajos de materia orgánica; 3) sostiene concentraciones muy bajas de bacterias del suelo; y 4) el suelo oxida igualmente L- y D-aminoácidos y L- y D-azúcares, debido a la presencia de oxidantes no biogénicos. Una de las similitudes más fascinantes entre los suelos del desierto de Atacama y los suelos de Marte, que también está vinculada a la sequedad extrema, es la acumulación natural de percloratos en la superficie del suelo. Los suelos de Atacama contienen la mayor concentración de perclorato encontrada en suelos de la Tierra y puede ser casi tan alta como las concentraciones observadas por la sonda Phoenix y el rover MSL en Marte. Además, el desierto de Atacama contiene densos (de decenas a cientos de metros) depósitos evaporíticos, que mantienen el

registro de antiguas condiciones acuosas y que son similares a los depósitos de sales identificados recientemente sobre regiones relativamente grandes de las tierras altas del sur de Marte (Figura 1).

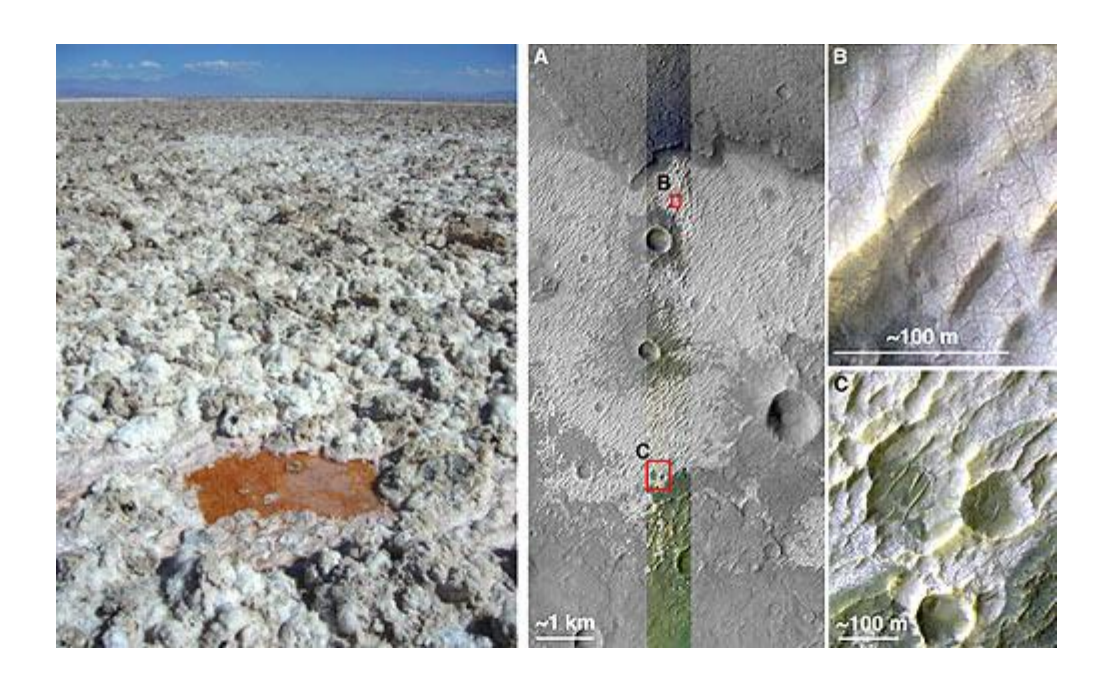


Figura 1: *Atacama: El Salar de Atacama (Chile), situado a los pies del volcán Licancabur en la precordillera andina, a 2500 metros sobre el nivel del mar y extendido sobre casi 100 km², conserva todavía algunas acumulaciones de agua entre los vastos depósitos de sales, como se puede ver en la imagen de la izquierda. El salar es producto del afloramiento de aguas subterráneas que, saturadas de sales, se evaporan dejando costras ricas en sal y minerales. La sonda Mars Odyssey ha revelado la presencia de más de 200 depósitos de sales en depresiones de hasta 25 km² distribuidas sobre las tierras altas del sur de Marte. El origen de estos depósitos puede encontrarse en un proceso análogo al que forma el Salar de Atacama. La imagen de la derecha muestra: (A) una vista general de algunos depósitos marcianos, y (B) y (C) aproximaciones en detalle que permiten apreciar la formación de polígonos por desecación. Los depósitos aparecen con tonalidades brillantes. (Fotografía de Atacama: A.G. Fairén; Marte: NASA/JPL/Arizona State University/University of Hawaii/University of Arizona).*

25.3.5. Análogo del Amazónico (II)

University Valley, en la Antártida. La Antártida está cubierta de hielo casi por completo, pero hay varios lugares libres de hielo que han sido reconocidos como análogos del regolito marciano. Los Valles Secos de la Antártida son la mayor región sin hielo del continente. En general, los valles tienen un clima desértico frío, la temperatura media anual oscila entre -17ºC y -25ºC, y la precipitación media anual es de menos de 50 mm/m². University Valley es un pequeño valle situado en las montañas Quartermain, que contiene un campo de nieve permanente en un extremo y un suelo de permafrost en el fondo del valle. Una comparación entre el perfil del hielo bajo la superficie en University Valley y el perfil observado en el subsuelo del polo norte de Marte por la sonda Phoenix ofrece similitudes interesantes, como que en ambos lugares

existe una capa de permafrost seco por encima del suelo de hielo macizo, que la cantidad de hielo en el suelo aumenta o disminuye debido al intercambio de vapor con la atmósfera (y no de agua líquida), y que las temperaturas en la parte superior del suelo de hielo pueden exceder los 273 K. Por lo tanto, el permafrost seco y el suelo de hielo de University Valley se pueden equiparar con los suelos de Marte. Además, el permafrost de University Valley no muestra signos de presencia de actividad microbiana que se pueda detectar con los métodos de los que disponemos actualmente: es el único lugar de nuestro planeta que la vida no ha podido colonizar.

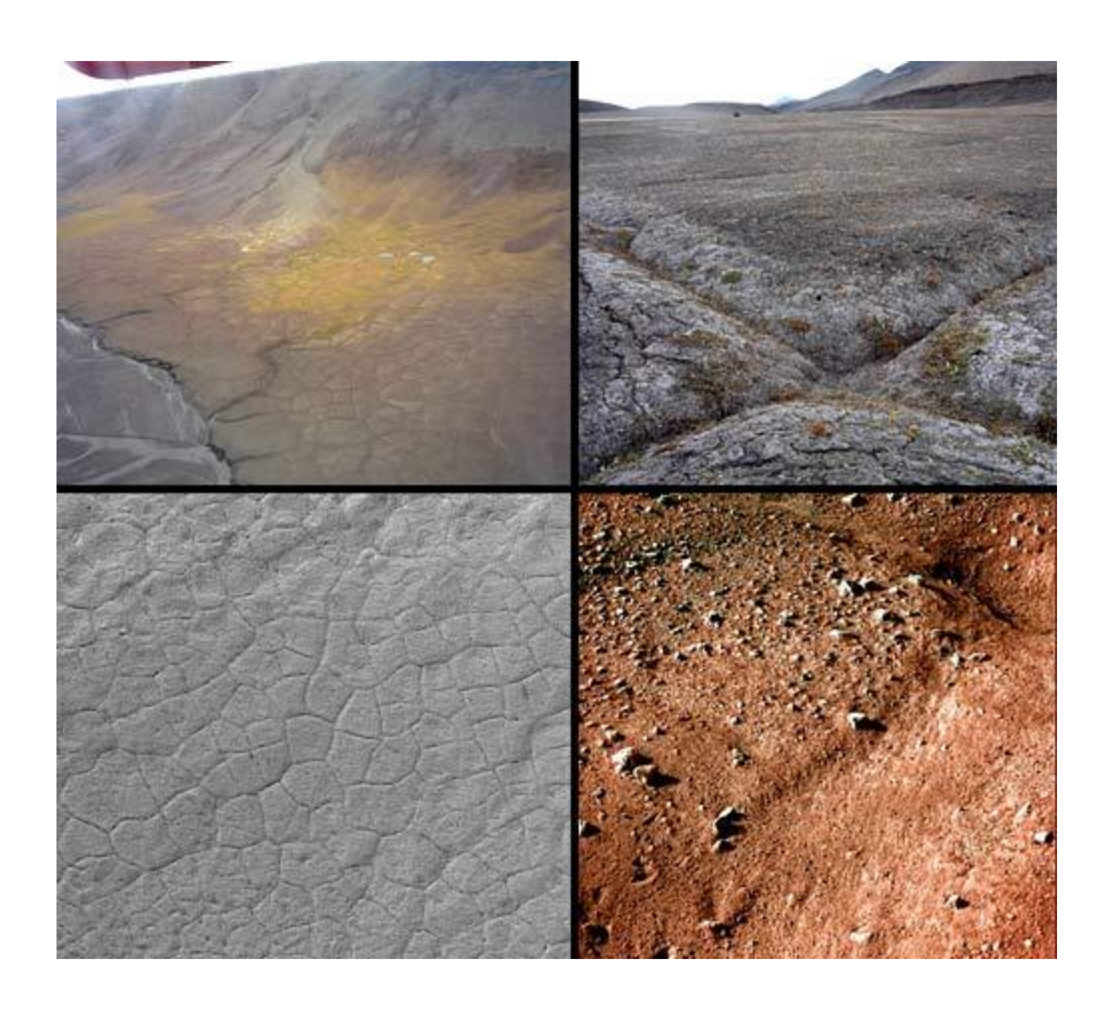


Figura 2: *Ártico: La fotografía superior izquierda muestra un terreno en forma de polígonos en la margen derecha del fiordo "Expedición", en la isla Axel Heiberg (Canadá), formados por la contracción y expansión estacional del hielo subsuperficial derivada de los cambios en la temperatura del suelo. Los polígonos miden entre 15 y 20 metros de lado. A la derecha, un detalle de las zanjas que configuran los polígonos. La imagen inferior izquierda fue obtenida por la sonda Mars Global Surveyor cerca del polo sur de Marte (86.9°S, 170.6°W), y muestra polígonos similares a los terrestres en morfología y tamaño. A la derecha, una imagen de la zona de amartizaje de la sonda Phoenix, que muestra polígonos de entre 1.5 y 2.5 metros de lado. Estos polígonos tan pequeños indican que el hielo no se encuentra a gran profundidad, y que se han producido cambios importantes en la temperatura en superficie recientemente. (Fotografías de Axel Heiberg Island: A.G. Fairén; Marte: NASA/JPL).*

25.4 Perspectivas

Las investigaciones de los análogos marcianos en la Tierra están proporcionando un mayor conocimiento sobre la evolución climática, geológica, geoquímica y mineralógica de la superficie marciana. Estas investigaciones mejorarán nuestra comprensión de la vida en ambientes extremos y la preservación de biomarcadores en la Tierra, así como de la posibilidad de la presencia y preservación de evidencias de vida extinta o presente en Marte.

Bibliografía:

Amils, R., González-Toril, E., Fernández-Remolar, D., Gómez, F., Aguilera, A., Rodríguez, N., Malki, M., García-Moyano, A., Fairén, A.G., de la Fuente, V., and Sanz, J.L. (2007). “Extreme environments as Mars terrestrial analogs: the Río Tinto case”. Planet Space Sci 55:370–381.

Andersen, D.T., Pollard, W.H., McKay, C.P., and Heldmann, J. (2002) “Cold springs in permafrost on Earth and Mars”. J Geophys Res 107, doi:10.10129/2000JE001436.

Benison, K.C. and La Clair, D.A. (2003) “Modern and ancient extremely acid saline deposits: terrestrial analogs for Martian environments?” Astrobiology 3:609–618.

Catling, D.C., Claire, M.W., Zahnle, K.J., Quinn, R.C., Clark, B.C., Hecht, M.H., and Kounaves, S. (2010) “Atmospheric origins of perchlorate on Mars and in the Atacama”. J Geophys Res 115, doi:10.1029/2009JE003425.

Mahaney, W.C., Dohm, J.M., Baker, V.R., Newsom, H.E., Malloch, D., Hancock, R.G.V., Campbell, I., Sheppard, D., and Milner, M.W. (2001) “Morphogenesis of Antarctic paleosols: martian analog”. Icarus 154:113–130.

Alberto González Fairén

Doctor en Biología Molecular

Investigador Centro de Astrobiología (CSIC-INTA) y Universidad Cornell.

Capítulo 26
¿Qué pasa con la basura espacial?

(Por Isabel Caballero de Frutos)

Desde el lanzamiento por la Unión Soviética del primer satélite artificial Sputnik 1, el 4 de octubre de 1957, la órbita terrestre ha sido colonizada incesantemente por satélites y objetos artificiales a medida que la tecnología espacial progresaba a marchas forzadas. A partir de ese momento, considerado como un hito en la historia de la Humanidad, la percepción del espacio y de nuestro planeta ha cambiado de manera notable. A lo largo de estas seis décadas, las aplicaciones de la teledetección han evolucionado desde el inicial carácter militar durante la Guerra Fría para desembocar en la actualidad en diversas actividades que abarcan las telecomunicaciones, el comercio, la seguridad, la navegación, la observación de la Tierra, la meteorología y oceanografía, y un largo etcétera de disciplinas que se nutren de este tipo de herramientas. Pese al extraordinario progreso científico sin precedentes que la exploración espacial ha generado en nuestras vidas cotidianas, con inmensos e incuestionables beneficios, nos encontramos con un inconveniente común. Todos los objetos artificiales que se encuentran orbitando alrededor de la Tierra tienen fecha de caducidad, a partir de la cual ya no darán servicio y por lo tanto, finalizará su "vida útil".

Esta realidad ha provocado que exista un número elevadísimo de objetos sin utilidad, fruto de la actividad del ser humano, lo que se conoce como "basura espacial" o también denominada "chatarra espacial" (Vídeo 1). La diversidad de desechos es enorme, englobando desde naves que se han quedado obsoletas, grandes restos de cohetes o artefactos desprendidos durante las misiones espaciales, hasta pequeñas partículas de pintura o polvo. Estos fragmentos espaciales se pueden clasificar según nuestra capacidad de detectarlos mediante radares y telescopios, siendo solo las fracciones mayores de 10cm las que se localizan de forma precisa. Según los expertos de las Agencias Espaciales Europea (ESA) y Americana (NASA), se estima que hoy en día solo un 7%, aproximadamente, de los objetos en órbita son naves operativas, que corresponden a alrededor de 1.000 satélites activos. Las cifras continúan, pues existen más de 150 millones de restos de tamaño inferior a un centímetro orbitando a nuestro alrededor. A primera vista, podríamos no dar importancia a este conjunto específico, pero teniendo en cuenta las elevadas velocidades a las que circulan estos fragmentos debido al campo gravitatorio (entre 6-10km/s), incluso los más diminutos suponen un potencial peligro, pudiendo actuar como proyectiles y dañar los satélites operativos. Asimismo, encontramos alrededor de medio millón de fragmentos entre 1-10cm y poco más de 20.000 piezas de más de 10cm. Cuando el tamaño es considerable, la amenaza es superior y el impacto puede ser desastroso, ocasionando deterioros irreparables. Recordemos la famosa colisión accidental en el espacio en el año 2009 del satélite estadounidense de comunicaciones Iridium-33 con el satélite militar ruso Kosmos-2251, la cual provocó la destrucción de ambos y generó más de 2.000 fragmentos de basura visibles. Incluso los astronautas, sin pretenderlo, han ensuciado el espacio inmediato a

nuestro planeta al perder durante sus paseos espaciales diversos objetos tan variopintos como guantes, bolígrafos, cámaras o herramientas.

Se ha publicado recientemente un vídeo-infografía en la que se muestra una reproducción de toda la basura espacial registrada en Space Track que rodea a nuestro planeta, repasando la historia entre 1957 y 2016 (Vídeo 2). Es simplemente una imagen general sin ajustarse a escala, pero ilustra perfectamente el colapso actual que existe en el espacio y nos plantea el inmenso desafío que esto provoca para la Humanidad.

Cabe mencionar que los deshechos orbitales no se encuentran distribuidos uniformemente, sino que hay algunas franjas especialmente transitadas o "sucias", donde se acumula la mayor densidad de basura. La órbita geoestacionaria GEO (Geostationary Earth Orbit), situada alrededor de los 36.000km de altura sobre la superficie terrestre, es una de las más utilizadas en el posicionamiento de satélites, debido principalmente a la estabilidad que presenta para este tipo de operaciones, siendo a la vez la más problemática. Esto es debido a que los objetos no escapan a la gravedad de la Tierra, pero tampoco son atraídos con suficiente fuerza como para desintegrarse, por lo que podrían residir ahí durante miles de años sin ningún control.

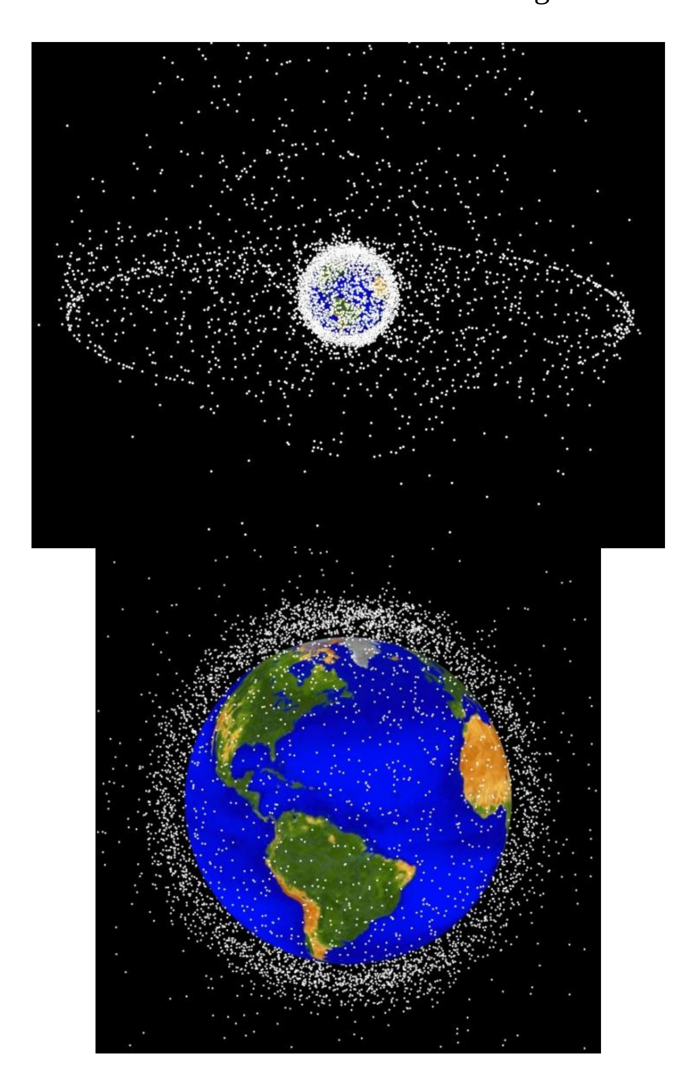


Figura 1: *Imágenes generadas por ordenador de los objetos en órbita terrestre que están siendo rastreados actualmente en la GEO (arriba) y LEO (abajo). Fuente: NASA Orbital Debris Program.*

Adicionalmente, las órbitas de baja altitud LEO (Low Earth Orbit) comprenden la zona situada entre 160 y 2.000km, y debido a la gravedad, los objetos localizados en ellas van descendiendo muy lentamente con el tiempo. Al encontrarse más próximas a la atmósfera terrestre, la cual es un protector efectivo, provoca que tras la entrada del deshecho, el rozamiento por fricción con los gases que la componen y la celeridad de caída fragmenten los objetos, con la desintegración y su destrucción en mayor o menor grado. Esta gran afluencia de residuos obsoletos ocasiona otro problema añadido: la generación de más basura debido a la colisión entre sí o el impacto con satélites en uso. Este es el efecto cascada o dominó que se conoce como Síndrome de Kessler. Donald J. Kessler, científico de la NASA en la segunda mitad del siglo XX, ya predijo el incremento exponencial de objetos espaciales alertando del peligroso volumen de basura, lo cual dificultaría la puesta en órbita de nuevos instrumentos.

La basura espacial apenas supone un peligro real para los seres humanos. El número de incidentes registrados como consecuencia de la caída de fragmentos en la superficie de la Tierra es relativamente reducido. Sin embargo, el mayor inconveniente reside en la colisión en las capas de la atmósfera. Mientras que los impactos con escombros de muy pequeño tamaño se producen con frecuencia sin ocasionar daños, son las colisiones de objetos con dimensiones superiores a los 10cm las que pueden originar consecuencias desastrosas, aunque las probabilidades sean mínimas. En la actualidad, las naves espaciales están capacitadas mediante blindajes robustos para resistir a impactos de piezas de hasta 1cm de tamaño. No obstante, como pronosticó Kessler, la chatarra espacial supone una amenaza directa para las misiones actuales y futuras o en las operaciones tripuladas, por lo que es necesario conocer y supervisar constantemente las zonas de mayor acumulación de deshechos a fin de evitar riesgos de colisión.

Las agencias espaciales y los diferentes gobiernos e instituciones ya trabajan de forma prioritaria desde hace años para remediar la acumulación de residuos, un reto para la tecnología aeroespacial debido a la complejidad y el coste económico que implica. En primer lugar, debemos mencionar las medidas que se encuentran dirigidas a mitigar o, preferiblemente, prevenir la producción de basura espacial. La Organización de Naciones Unidas cuenta con una normativa por la que los países deben comprometerse y presentar un plan de mitigación con la puesta en órbita de instrumentos espaciales. Sin embargo, no existe un marco legal internacional que obligue a los Estados a limitar o limpiar la producción de deshechos en el espacio, por lo que parece necesario reforzar la legislación. Varias agencias, como la NASA y la ESA, ya han establecido algunas normativas internas, con directrices y líneas de actuación recomendadas para su cumplimiento. Se pueden citar varios ejemplos, como son evitar la destrucción intencionada de objetos, prevenir las explosiones accidentales de las fases operativas o reducir los residuos derivados de las mismas con los sistemas de desecho. Igualmente, otra de las medidas consiste en limitar la presencia de satélites y propulsores a un número de años en las zonas más saturadas, como lo son la banda de los 800km en la LEO y toda la GEO, para que entren en la atmósfera o sean almacenados en órbitas más alejadas denominadas “órbita cementerio”, donde no supongan peligro para los equipos activos.

Pero el sólido consenso entre los expertos en la materia es que, con el escenario expuesto, la futura protección y preservación del espacio requerirá de otro tipo de actuaciones. Será necesario combinar las medidas preventivas con las estrategias de

retirada activa y limpieza de la basura espacial existente, lo que entraña un desafío de enorme envergadura. La investigación y desarrollo de estas misiones debe ser abordada como un esfuerzo conjunto de cooperación entre industria, gobiernos y agencias espaciales internacionales. Las soluciones que están sugiriendo los científicos y expertos de todo el mundo son muy diversas, trabajando para combatir el problema de la manera más eficaz y económica.

En este sentido, la ESA dispone de un programa pionero, la Iniciativa Espacio Limpio "Clean Space", dividido en varias categorías metodológicas encargadas de reducir el impacto ambiental de la industria espacial tanto en la órbita como en tierra. Por un lado, el proyecto se centra en el diseño para fomentar las tecnologías ecológicas (EcoDesign) y reducir la producción de desechos espaciales (CleanSat), como también en la investigación de productos no contaminantes en la fabricación de satélites, desde el combustible hasta las materias primas. Por otro lado, con la iniciativa "e.Deorbit", se estudia la recuperación activa de artefactos en la órbita LEO. El objetivo de la misión es utilizar una nave espacial para capturar los objetos en desuso propiedad de la ESA, acercándolos hacia la Tierra de forma controlada y desintegrándose de manera segura a lo largo del proceso. En la actualidad hay dos conceptos bajo consideración: uno con una red y el otro con un brazo robótico.

Otro experimento propuesto por la Agencia Espacial Japonesa (JAXA) es el uso de una gigantesca red de un material muy resistente para capturar los restos de sondas y satélites que van flotando por el espacio. Lo que pretenden crear es una red hecha de hilo especial, capaz de resistir cambios extremos de temperatura y grandes fuerzas, que tras ser lanzada junto con un satélite y desplegada, cubra varios kilómetros. A continuación, comenzará a orbitar la atmósfera terrestre recolectando los deshechos.

La NASA se encuentra experimentando sobre un posible gel aéreo, un material sintético y ligero que se solidifica y provocaría que la basura se pegue a él. Esta agencia ya hizo uso de este gel para recolectar polvo espacial en la nave Stardust, pero aún no se ha probado con basura espacial. Otra idea explorada desde hace tiempo por los EEUU y China es el uso de un sistema de rayos láser poco potentes con el que se vaporizaría la basura en pequeños fragmentos para reducir la altura de su órbita y, de este modo, arrastrarlo lentamente hacia la Tierra.

En España, investigadores de la Universidad Politécnica de Madrid han liderado un proyecto internacional BETs (The Bare Electrodynamic Tethers Project) donde plantean el desorbitado de la basura espacial en la órbita LEO mediante amarras electrodinámicas. Esta tecnología consiste en incluir una amarra enrollada como un carrete en el satélite, y cuando este deje de estar operativo, la amarra se desplegaría de forma pasiva, utilizando el frenado del campo geomagnético en lugar del débil frenado atmosférico. Entre sus prometedores resultados está el desarrollo de un simulador de utilidad para agencias espaciales y empresas.

La misión RemoveDebris, dirigida por el Centro Espacial de Surrey, es un primer intento de la UE [1] destinado a recoger basura espacial, la cual tiene previsto comenzar en 2017. Se trata de una misión experimental que probará tres sistemas para capturar o desviar los deshechos espaciales con tecnología a bajo coste, empleando instrumentos como un arpón, una red y un sistema de navegación para arrastrar de forma remota a los satélites viejos a lugares donde se desintegren.

Existen otros proyectos de mayor complejidad que proponen diferentes soluciones a este problema, como la puesta en órbita de un vehículo espacial que, mediante brazos articulados, permita recoger los desechos espaciales de gran tamaño. Éstas y otras ideas ya planteadas son imprescindibles para abordar la búsqueda de un método eficaz, económico y ecológico, el objetivo principal de investigadores y científicos especializados en el sector espacial. Pero, por el momento, un gran porcentaje de estos diseños supera nuestra capacidad técnica y nuestros recursos económicos.

Mientras se resuelven estos aspectos asociados a los incentivos del problema, es fundamental que la comunidad internacional sea consciente de la importancia de la proliferación de objetos en órbita y siga avanzando en su exploración, tomando algunas decisiones importantes.

La vulnerabilidad tanto de la Tierra, como del espacio que la rodea, saturada de desechos generados por los humanos, es una llamada de atención imperiosa para la búsqueda de futuras soluciones que, además, ha pasado de pertenecer exclusivamente al ámbito de la gestión aeroespacial para irrumpir con auge en la opinión pública internacional.

Por un Universo limpio como herencia para las generaciones venideras.

Vídeo:
1: http://www.esa.int/spaceinvideos/Videos/2013/04/Space_debris_story
2:https://www.youtube.com/watch?annotation_id=annotation_3598026871&feature=iv&src_vid=wPXCk85wMSQ&v=O64KM4GuRPk

Notas:
1. https://directory.eoportal.org/web/eoportal/satellite-missions/r/removedebris

Bibliografía:
Heiner Klinkrad, Space Debris - Models and Risk Analysis. Springer Verlag, 2006, 430, ISBN 3-540-25448-X

Isabel Caballero de Frutos
Doctora en Ciencias
Investigadora postdoctoral
Instituto de Ciencias Marinas de Andalucía (ICMAN-CSIC)

Capítulo 27
¿Por qué unos suelos resisten a la erosión y otros no? La estructura del suelo

(Por José Luis Rubio)

La respuesta se encuentra en la más fascinante interfaz del mundo mineral y el mundo biológico de nuestro planeta. En las profundidades del suelo y descendiendo hasta escalas microscópicas pueden desarrollarse uniones íntimas, complejas y perdurables entre las partículas inorgánicas procedentes del mundo geológico y los componentes orgánicos provenientes del mundo vegetal. Esta frontera del mundo mineral y del mundo biológico origina una propiedad clave que llamamos estructura del suelo y que, entre otras muchas funciones, es responsable de la resistencia del suelo a la erosión.

Pero vayamos por partes. Existen zonas en las que la erosión se muestra de forma intensa y bien visible, a veces con paisajes desgarrados e intensamente afectados por procesos erosivos que impactan en la disposición natural del suelo provocando su desmantelamiento, desagregación y arrastre. Si el terreno se encuentra en pendiente, se mostrará profundamente hendido y rasgado por una impactante red de surcos, arroyos, cárcavas y barrancos que muestran una tierra herida y degradada en la que el suelo fértil superficial va desapareciendo por el efecto inmisericorde de aguaceros y torrentes. Sin embargo, junto a ellas y bajo las mismas condiciones climáticas quizás puedan observarse terrenos con suelos fértiles y profundos que mantienen una abundante cobertura de vegetación y que dan lugar a paisajes estables y productivos.

¿Por qué estas grandes diferencias en el comportamiento del suelo frente a la erosión?

Conviene indicar que en los procesos erosivos existen factores relacionados con la propia naturaleza del suelo, y otros factores externos que también influyen en el proceso. Brevemente, estos últimos incluirían las condiciones climáticas (incluyendo la precipitación, temperatura, evapotranspiración, torrencialidad y erosividad de la lluvia, periodos de sequía, etc.), los factores topográficos (ángulo de la pendiente, su tipología y longitud), la naturaleza de los materiales geológicos originarios del suelo, el tipo de uso del suelo (agrícola, forestal, pastoral, abandonado, urbano, afectado por infraestructuras,…) y el manejo concreto de cada suelo, incluyendo el laboreo, las medidas de conservación, la fertilización, la gestión y manejo del agua, etc.…

Indudablemente todo este amplio conjunto de circunstancias extrínsecas ejerce una gran influencia sobre la intensidad de los procesos erosivos pero en este capítulo nos centramos en un factor interno que puede representar una espectacular diferencia en la estabilidad y resistencia a la degradación entre distintos tipos de suelos.

Hay que señalar que en la respuesta a la erosión existen numerosas circunstancias físicas, químicas y biológicas que derivan de la complejidad del suelo como interfaz geológica, hídrica, atmosférica y biológica. Su papel como membrana viva que envuelve al Planeta implica una manifiesta complejidad que hace casi imposible identificar una sola propiedad para explicar un determinado comportamiento. Sin embargo si tuviéramos que elegir un solo rasgo para explicar la respuesta erosiva de un suelo este sería la estructura del suelo o estructura edáfica.

Aparentemente la estructura del suelo es algo simple: es la disposición entre sí de las distintas partículas orgánicas e inorgánicas del suelo. Pero esta simple definición encierra un insospechado universo microscópico. Lo que podemos observar a simple vista de ella son las unidades de agregados o terrones del suelo que se van formando por la adherencia de otras pequeñas subunidades de agregados. En ellos podremos observar, por ejemplo, el papel mecánico de ligazón de las hifas de hongos. Para ir desentrañando el mecanismo de agregación hay que ir descendiendo al nivel de micro agregados de menos de 1mm que ya no pueden ser observados a simple vista. Pero para conocer la esencia del mecanismo interno todavía es necesario descender en la escala de observación a niveles de menos de 0,1mm, y mucho más, pasando a niveles de milimicras y luego a dimensiones moleculares que lógicamente requieren sofisticados equipos de observación.

En este submundo microscópico podremos encontrar la respuesta a la pregunta de por qué después de millones y millones de años de lluvias de todo tipo no se ha ido lavando y desapareciendo la piel viva del planeta que representan los distintos tipos de suelo sobre su superficie. La base de la respuesta se encuentra en el juego al que permanentemente se dedican las cargas iónicas de las partículas coloidales de la tierra. En los niveles moleculares del suelo se llega a la frontera del mundo mineral y del orgánico, que comentábamos al principio, representado por la interfaz entre macromoléculas de humus y la superficie de partículas minerales como son las arcillas. En esta interfaz las arcillas tienen preferentemente cargas negativas y las macromoléculas húmicas también. En principio cargas del mismo tipo se repelerían e impedirían la unión entre partículas, pero para evitarlo, entre humus y arcillas se disponen cationes cargados positivamente, tales como calcio, magnesio, amonio o potasio, que actúan como enlaces. Este mecanismo que solo hemos esbozado (existen distintos tipos de arcillas, diferentes fuerzas de unión electroquímica, distintos cementantes y una enorme variedad de compuestos húmicos) es la base de la resiliencia del suelo y de numerosas funciones y propiedades del mismo. Lo que percibimos a simple vista, y sus implicaciones en el paisaje natural o en uso del suelo, es la traslación de lo que existe a escala molecular en las interacciones electroquímicas entre partículas minerales y orgánicas.

La estructura edáfica no es algo estable y definitivo. Es una propiedad dinámica que puede mejorarse o que puede empeorarse por maltrato del suelo o por cambios en las condiciones climáticas. Si mejoramos y conservamos la estructura edáfica todo serán ventajas. El suelo será más resistente a los procesos erosivos. Infiltrará y retendrá mayor cantidad de agua con lo cual soportara mejor la falta de precipitaciones y sequías. La tierra estará más aireada con lo cual se activara su potencial biológico y los procesos enzimáticos que aumentan los nutrientes. El suelo será más estable, más fácil de trabajar y, en una palabra, será un medio más favorable como hábitat para las plantas y sus raíces. Es decir será más fértil y productivo. Podríamos seguir con una larga lista de ventajas pero en la situación actual de cambio climático, no podemos olvidar que una

buena estructura edáfica también potencia la capacidad del suelo como sumidero y depósito de CO_2 atmosférico, es decir contribuye significativamente a la mitigación del cambio climático.

Podemos concluir que el que se produzca o no esa feliz sintonía entre las fronteras moleculares del mundo orgánico e inorgánico, que da lugar a una buena estructura, marcará la diferencia entre suelos y paisajes ricos y fértiles o suelos erosionados, degradados e infértiles.

Platon en el Citrias nos ofrece una descripción del Ática que nos sirve como ejemplo de intensa erosión a evitar: “la tierra que era la mejor del mundo, se parece ahora a un cuerpo demacrado por la enfermedad en el que se escurre por todas partes la tierra vegetal y fecunda y solo nos queda un cuerpo descarnado hasta los huesos”. Los atenienses, como otras muchas culturas, no cuidaron adecuadamente de la estructura del suelo.

José Luis Rubio
Doctor Ingeniero Agrónomo
Investigador Científico CIDE – CSIC, Valencia

Capítulo 28
¿Podrían los océanos volverse tan salados como para matar toda la vida marina?
(Por Joaquim Ballabrera Poy y Emilio García Ladona)

Alrededor del 71% de la superficie de la Tierra está cubierta por mares y océanos. La principal característica del agua marina es su sabor salado. La salinidad de los océanos se refiere a la cantidad de sales disueltas en el agua de mar. La "salinidad" del agua dulce de la mayoría de ríos y lagos es por lo general alrededor de un gramo, o menos, de sales disueltas por cada kilogramo de agua. La concentración promedio de la salinidad de los océanos es de unos 35 gramos de sales en cada kilogramo de agua de mar. En los estuarios, donde el agua de mar se mezcla con el agua dulce de los ríos, el rango de salinidad varía entre los dos valores mencionados. Las aguas con mayor concentración de sales del mundo se encuentran en el lago San Juan, en la Antártida, donde la concentración de salinidad puede llegar a 440 gramos de sal por kilogramo de agua. En este tipo de ecosistemas hipersalinos viven únicamente especies específicas de crustáceos y microbios. El volumen total de los océanos es de aproximadamente $1,35 \times 10^9$ km^3 = $1,35 \times 10^{18}$ m^3. Un sencillo cálculo nos proporciona una estimación aproximada de la cantidad total de sales en el océano. Si la densidad media de los océanos es 1.027 kg / m^3, la masa total de los océanos es de aproximadamente $1,39 \times 10^{21}$ kg, y la cantidad total de sales es, pues, de $4,85 \times 10^{16}$ toneladas.

Desde mediados del siglo XVIII, los químicos han desarrollado métodos para el análisis químico del agua, desarrollados originalmente para ayudar a purificar el agua potable. Hoy sabemos que el agua de mar es una solución de todos los elementos presentes en la naturaleza. En mar abierto, los iones disueltos más abundantes son el cloruro (alrededor de 19,0 gramos), sodio (10,6 gramos), sulfato (2,7 gramos), magnesio (1,3 gramos), calcio (0,4 gramos), potasio (0,4 gramos), bicarbonato (0,1 gramos), bromuro, ácido bórico (menos de 0,1 gramos), y estroncio y fluoruro (menos de 0,01 gramos por kilogramo de agua de mar). A partir del análisis químico de las muestras recogidas durante la expedición del H.M.S. Challenger (1872-1876), se encontró que la relación de las concentraciones de dichos componentes es muy similar en la mayor parte del océano abierto. Estos componentes se denominan conservativos debido a que sus concentraciones no se ven afectadas por las reacciones biológicas y químicas que comúnmente tienen lugar en los océanos. La mayor parte del resto de los componentes del agua son llamados no conservativos debido a que sus concentraciones cambian de manera significativa debido a la actividad biológica y reacciones químicas presentes en los océanos.

28.1 Procesos que añaden y eliminan la sal del agua marina

Para explicar por qué el mar es salado y qué sales contiene, rápidamente se hizo hincapié en el hecho que el agua es un potente disolvente donde los azúcares y sales

inorgánicas se disuelven fácilmente. En los continentes, el agua de los ríos disuelve las rocas y sedimentos (meteorización) en su camino hacia el mar. Ese proceso ya se había mencionado en el siglo XVI como el origen de la salinidad de los océanos. De hecho, cada año los ríos vierten cerca de cuatro mil millones de toneladas (4×10^{12} kg) de sales disueltas en el océano. Si la salinidad del océano se debiera únicamente a la erosión de las rocas de la corteza continental, la salinidad del océano tendría que aumentar con el tiempo debido al continuo fluir de los sedimentos desde los continentes hacia los océanos. Por otro lado, la composición química relativa de ríos y mares debería ser similar (aunque los mares tengan una concentración más alta). Pero hoy sabemos que éste no es el caso: la composición relativa del agua marina y de los ríos es significativamente diferente (Tabla 1). En efecto, si la erosión de las rocas de la corteza continental fuera el único origen de sales de agua de mar, la composición del agua de mar debería ser más rica en magnesio y más pobre en cloruro y calcio.

	AGUA MARINA		*AGUA DE RÍO*	
SUSTANCIA	*Concentración*	*Relativaa* Cl^-	*Concentración*	*Relativaa* Cl^-
Bicarbonato (HCO_3^-)	0,105	0,0055	0,0530	6,3855
Calcio (Ca^{+2})	0,404	0,0213	0,0147	1,7711
Chloruro (Cl^-)	18,980	1,0000	0,0083	1,0000
Magnesio (Mg^{+2})	1,272	0,0670	0,0037	0,4458
Potasio (K^+)	0,390	0,0205	0,0014	0,1687
Sílice (Si)	0,003	0,0002	0,0049	0,5904
Sodio (Na^+)	10,556	0,5562	0,0072	0,8675
Sulfato (SO_4^{-2})	2,649	0,1396	0,0115	1,3855

Tabla 1. *Concentraciones típicas (gramos por kilogramo) de los principales constituyentes del agua de mares y rios. Unidades en %. A partir de la Encyclopedia of Global Change: Environmental Change and Human Society, Volume 1, Edited by Andrew S. Goudie and David J. Cuff, Oxford University Press, 2001. 669 pp.*

Uno de los procesos que ayudan a explicar la diferente composición de ríos y mares es la actividad volcánica. En efecto, las erupciones volcánicas son ricas en compuestos de cloro, carbono y azufre. Los gases solubles en agua expulsados por los volcanes se disuelven en la lluvia y precipitan. La deposición de aerosoles también aporta compuestos de origen volcánico a los océanos. De hecho, cuando se compara la composición química del agua de mar con la composición química de la corteza continental y la del manto superior terrestre, es esta última la que presenta una mayor similitud.

El último proceso que juega un papel clave en la composición química del agua de mar fue, descubierto a finales de la década de 1970, con el descubrimiento de los respiraderos hidrotermales de las profundidades marinas, que se encuentran comúnmente en el fondo del mar cerca de lugares de actividad volcánica y a lo largo de las dorsales oceánicas. Erupciones en estos respiraderos eyectan columnas de agua

sobrecalentada rica en minerales (Ilustración 1). La temperatura del agua que emana de estos orificios de ventilación varía de 25°C a 400°C. En estas zonas, el agua de los océanos puede circular a través de una corteza oceánica rota y altamente porosa, permitiendo el contacto entre rocas fundidas y el agua de mar. Las altas temperaturas y la presiones permiten el establecimiento de una serie de reacciones químicas que modifican la composición química tanto del agua marina como de la corteza oceánica. En particular, el agua de mar pierde magnesio y gana calcio. Se estima que la totalidad del agua oceánica circula a través de dichas regiones hidrotermales en tan solo uno o dos millones de años.

Por otra parte, diversos procesos biológicos y físicos eliminan las sales del agua marina. Por ejemplo, algunas especies de fitoplancton (diatomeas) y cocolitóforos eliminan el calcio y silicio del agua cerca de la superficie del océano. Estas especies usan estos elementos para formar sus conchas. Las pesadas partículas de carbonato de calcio ($CaCO_3$), sílice particulado (SiO_2) y pirita (FeS_2) acaban por precipitar hacia el fondo del océano, donde son enterradas y eliminadas del sistema. Por otra parte, el carbono eliminado por el fitoplancton para construir sus tejidos blandos acaba por convertirse en alimento para los organismos superiores y bacterias.

Otro proceso por el cual el océano pierde sales es el de la evaporación, que puede conllevar la precipitación de minerales, especialmente de cloruro de sodio (NaCl) y sulfato de calcio ($CaSO_4$) por la desecación de masas aisladas de agua. Un ejemplo es el de la Crisis Salina del Messiniense, cuando el cierre del estrecho de Gibraltar permitió que el mar Mediterráneo se desecara a lo largo de un proceso cíclico de desecación-inundaciones que puede haber durado 630.000 años. Cuando el estrecho de Gibraltar volvió a abrir por última vez, los sedimentos habían enterrado a una cantidad estimada de $0,4 \times 10^{16}$ toneladas de sales.

Cambios en la solubilidad de las aguas oceánicas debido a cambios en la alcalinidad provocan la precipitación de algunas de las sustancias disueltas. Por ejemplo, las concentraciones de magnesio y calcio se encuentran en un delicado equilibrio y cualquier pequeño cambio en la alcalinidad conduce a su precipitación. Nódulos de manganeso se pueden encontrar en el sedimento del fondo marino en la mayoría de los océanos. Además, el metabolismo bacteriano en los sedimentos modifica continuamente la química del agua de mar, alternado la solubilidad de los materiales que se encuentran en los sedimentos.

La pulverización de agua marina, consistente en partículas expulsadas de la superficie del mar por el estallido de burbujas (como por ejemplo las producidas cuando rompen las olas), elimina iones disueltos en el agua de mar. Cloruro sódico e iones tales como K^+, Mg^{+2}, Ca^{+2} y SO_4^{-2} se eliminan del agua marina mediante este proceso. Por cierto, el proceso de pulverización también juega un papel fundamental en el ciclo del azufre al favorecer la emisión de sulfuro de dimetil (DMS), el principal gas biogénico emitido desde el mar, responsable del olor a mar, y que influye en la formación de nubes.

28.2 Tiempos de residencia

Un esquema de los principales procesos de la adición y la eliminación de los iones disueltos en el agua marina se presenta en la Ilustración 2. Existen evidencias que

indican que la concentración de la sal del mar ha cambiado poco en los últimos 1,5 mil millones de años. Por otra parte, la tolerancia a la salinidad de las bacterias que vivieron hace 3,8 mil millones de años indica que la salinidad ya habría sido similar en ese momento. La estabilidad en los niveles de salinidad se ve favorecida por el hecho de que la mayor parte de los mecanismos de eliminación de sales del agua de mar están directamente relacionados con su concentración. A mayor concentración, mayor tasa de eliminación de sal. En la década de 1950, los geólogos desarrollaron el concepto océano estacionario. En los sistemas estacionarios, la velocidad de adición de cualquier propiedad debe ser igual a la velocidad con la que dicha propiedad es eliminada. En tal caso, es posible estimar el tiempo que una partícula o elemento tarda en abandonar el sistema (el tiempo de residencia). Para un elemento dado en un estado de equilibrio, el tiempo de residencia se define como:

$$TiempodeResidencia = \frac{Concentración}{Tasadeentradaodesalida}$$

El tiempo de residencia de una sustancia depende de su reactividad química. Por ejemplo, los iones de hierro y aluminio son altamente reactivos y permanecen en el agua de mar durante un corto período de tiempo antes de ser incorporados en los sedimentos. En el otro extremo del espectro, los iones de cloruro permanecen en el agua de mar durante millones de años antes de ser retirados del océano. Estimaciones de los tiempos de residencia de los principales constituyentes del océano se dan en la Tabla 2.

SUSTANCIA	*TIEMPO DE RESIDENCIA*
Bicarbonato (HCO_3^-)	110.000
Calcio (Ca^{+2})	1.000.000
Cloruro (Cl^-)	100.000.000
Magnesio (Mg^{+2})	13.000.000
Potasio (K^+)	12.000.000
Sílice (Si)	20.000
Sodio (Na^+)	68.000.000
Sulfato (SO_4^{-2})	11.000.000

Tabla 2. *Estimación de los tiempos de residencia de los principales constituyentes del agua marina. Expresados en años. A partir de Wallace Broecker and Tsung-Hung Peng, Tracers in the Sea. A publication of the Lamont-Doherty Geological Observatory, Columbia University, New York, 1982. 690 pp.*

En conclusión, y en respuesta a la pregunta en el título de este capítulo, los elevados tiempos de residencia de los principales constituyentes de la salinidad del agua marina impiden cualquier cambio rápido de las concentraciones globales de salinidad si no se produce ningún gran cataclismo geológico en los próximos cien millones de años.

Bibliografia:
"Oceanography: an invitation to Marine Science", Tom Garrison and Robert Ellis. Cengage Learning, Boston, MA, USA, 2015. 640 pp.

Joaquim Ballabrera Poy **Emilio García Ladona**
Doctores en Ciencias Físicas
Dept. de Oceanografía Física y Tecnológica
Instituto de Ciencias del Mar (ICM-CSIC)

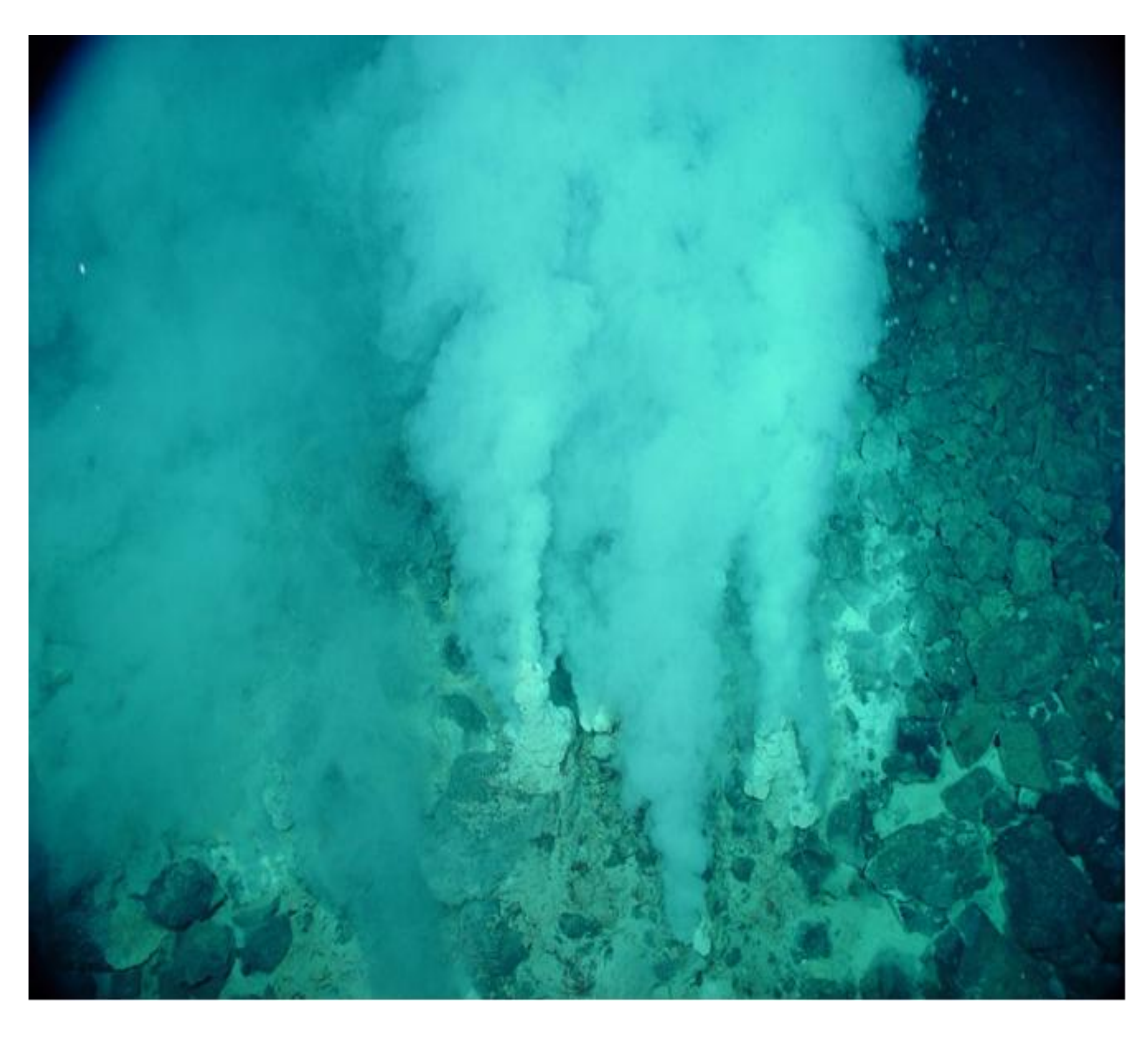


Illustration 1: *White chimneys at Champagne Vent site, NW Eifuku volcano (Japan). The chimneys are about 20 cm across and 50 cm high, venting fluids at 103ºC. Source* *http://oceanexplorer.noaa.gov/explorations/04fire/logs/hirez/champagne_vent_hirez.jpg*

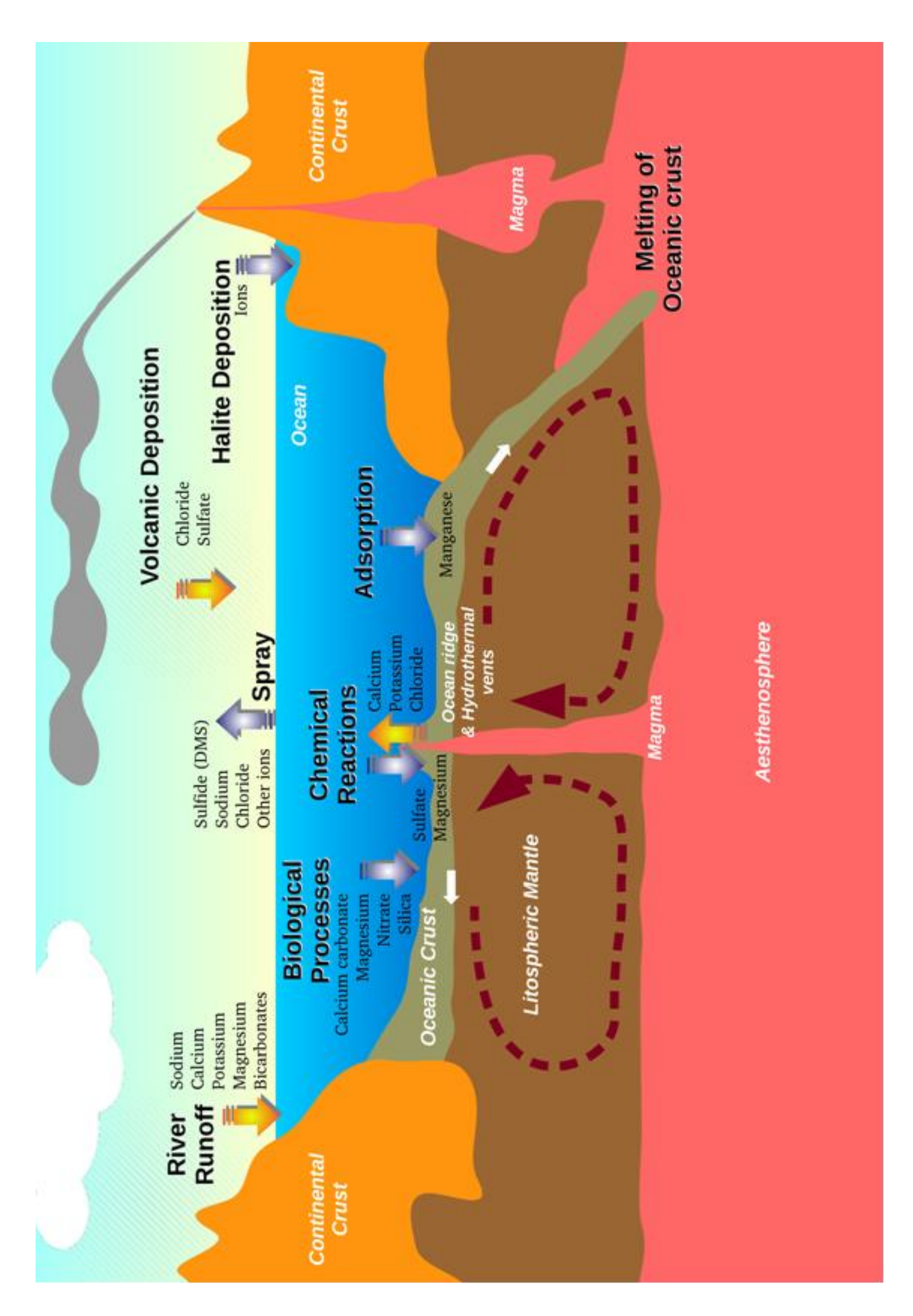


Illustration 2: *Schematic cross-section illustrating the main processes adding (orange arrows) and removing (blue arrows) ions from sea water.*

Capítulo 29

¿Qué ocurriría si se derritieran los casquetes polares? ¿El progresivo deshielo es achacable a la acción humana?

(Por Emilio García Ladona y Joaquim Ballabrera Poy)

Recurrentemente oímos noticias sobre el aumento del nivel del mar y la amenaza que ello supone para una población mundial cada vez más localizada en las zonas litorales, especialmente para los doscientos millones de humanos que viven a menos de 5 m de altura respecto el nivel del mar. ¿Pero qué procesos influencian en el nivel del mar? ¿Y qué sabemos hoy en día sobre la tendencia futura?

Los procesos que determinan el nivel del mar en cada región son múltiples y actúan en diferentes escalas de tiempo. El agua está repartida alrededor de diferentes compartimentos. En la atmósfera en forma de gas y pequeñas gotas en las nubes; en la litosfera en aguas freáticas, ríos, lagos, acuíferos incluyendo reservorios artificiales de origen antropogénico; en la criosfera, en las capas de hielo de la Antártida, del Ártico y Groenlandia y glaciares continentales. Sin embargo los reservorios de agua por excelencia son los mares y océanos. Así en gran medida la cantidad de agua en las cuencas oceánicas determina el nivel medio del mar. Hay dos formas de medir el nivel del mar. Una consistiría en la medición del grosor de la columna de agua en un punto, tal y como se ha realizado históricamente con ayuda de mareógrafos en las zonas costeras. Sin embargo, a partir del lanzamiento de los primeros satélites dotados de altímetros (i.e. TOPEX-Poseidon) se puede medir la variación de la superficie del mar en referencia a un nivel de equilibrio asociado al campo gravitatorio terrestre. Es decir aquella superficie cuyos puntos tienen la misma energía potencial gravitatoria. Si la Tierra fuera perfectamente redonda esta superficie sería también una esfera alrededor de la Tierra. Imaginemos una cebolla esférica, una superficie equipotencial sería una capa de esa cebolla a una distancia determinada del centro. Cada capa de la cebolla sería una superficie equipotencial y el nivel del mar tendría la misma forma. Sin embargo, la Tierra ni es redonda, ni su composición de materiales es homogénea y rota alrededor de su eje, por lo que las superficies equipotenciales distan mucho de ser esféricas. Las reconstrucciones más precisas que se han realizado hasta ahora se han obtenido a través de la misión GOCE de la Agencia Espacial Europea en 2011 (ver figura 1). Se pueden apreciar separaciones de hasta 200 m de altura entre diferentes áreas del globo. Para entendernos, un objeto colocado al Sur de Ceilán tiene la misma energía potencial que un objeto situado al Oeste de Irlanda, pero entre ellos hay una diferencia de altura de unos 150 m. El nivel del mar como respuesta a este campo gravitatorio debiera tener una forma aproximada al geoide. Sin embargo hay muchos otros factores que hacen que el nivel del mar instantáneo se separe del geoide. Y estos procesos tienen una variabilidad característica que produce variaciones apreciables del nivel del mar a escalas de tiempo que abarcan desde las pocas horas hasta los millones de años.

Uno de los factores más evidentes que contribuyen a la variación del nivel del mar es el aporte de agua a las cuencas oceánicas que tiene que ver básicamente con la

conexión de los diferentes reservorios de agua: descargas de ríos proveniente de precipitaciones y deshielo, precipitaciones directas sobre el mar y deshielo de casquetes polares de hielo. Sin embargo, las aportaciones más relevantes son los glaciares continentales que abocan masas de hielo directamente al mar y las capas de hielo continental en la Antártida y Groenlandia básicamente. A menudo, todavía hay gente que asocia el deshielo del Ártico con las variaciones del nivel del mar, pero eso es rotundamente erróneo. Una masa de hielo en agua retiene el mismo volumen total de agua debido al principio de Arquímedes. Así solamente casquetes de hielo sobre tierra firme tienen una contribución neta a ello.

El último informe de IPCC Panel (AR5) completado en el 2014 es un buen texto de referencia para conocer el estado del arte del conocimiento que tenemos sobre el nivel del mar y de los distintos procesos causantes de su evolución. Las reconstrucciones que disponemos indican un aumento relativamente "rápido" del nivel del mar desde el último máximo glacial (~20.000 a.c), unos 125 m hasta el valor actual que se establece alrededor de 2000 a.c. Al final del periodo conocido como Pequeña Edad de Hielo, ("Little Ice Age") entre mitad del siglo XIV y mediados del XIX, durante el cual el nivel del mar había bajado unos 15-18 cm, el nivel del mar empezó a aumentar progresivamente hasta nuestros días, unos 18-20 cm por encima del periodo pre-industrial (ver figura 1). Las últimas reconstrucciones estiman que la contribución de los hielos en Groenlandia y Antártida es del orden de 0.6 mm/año durante el periodo 1993-2010 que es un periodo donde se tiene mejores medidas gracias a una mejora de los programas de observación.

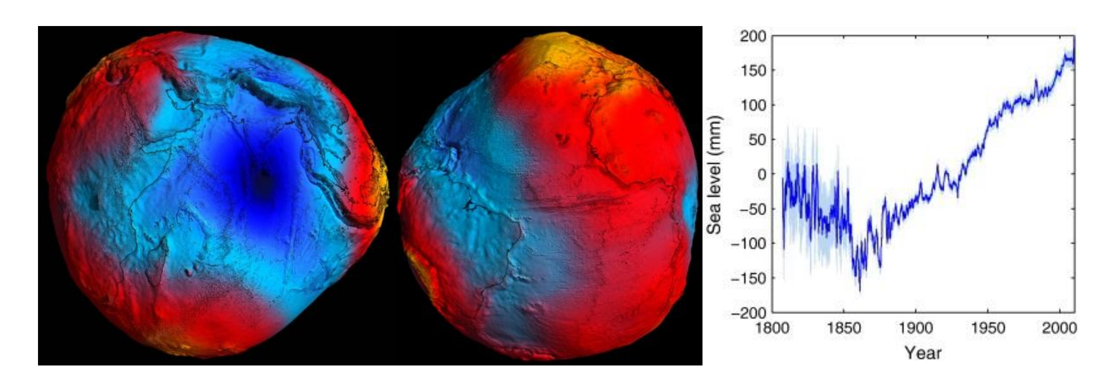


Figura 1: *Izquierda: Reconstrucción del geoide terrestre a partir de los datos del satélite GOCE de la ESA (© ESA). Derecha: Reconstrucción del nivel del mar entre 1807-2009 a partir de mareógrafos (reproducido a partir de Jevrejeva et al., 2014).*

La dinámica del hielo de los casquetes polares es muy compleja y depende del balance entre la acumulación de nieve, el deshielo (mayormente superficial) y el flujo de hielo saliente sobre el mar o directamente por desprendimiento directo y brusco de icebergs (“ice calving”). En el caso de Groenlandia las pérdidas de masa de hielo se redistribuyen aproximadamente por igual entre el deshielo en superficie y el desprendimiento sobre el mar. En el caso de la Antártida las temperaturas son muy bajas y básicamente el hielo se pierde fluyendo hacia la plataforma continental donde forma una cobertura de hielo que posteriormente va desprendiéndose en sus partes más externas y se va fundiendo por debajo en contacto con el agua. Estas plataformas de hielo apenas contribuyen al nivel del mar ya que están flotando sobre el agua sin embargo pueden indirectamente acelerar o frenar el flujo de hielo desde las zonas de acumulación como intentaremos ilustrar más adelante.

Para la cobertura de hielo en Groenlandia se da la paradoja que ha habido un aumento de acumulación de nieve en la zona interior pero también ha aumentado el deshielo superficial y el desprendimiento de manera que el balance neto indica una pérdida de la masa de hielo. En este último caso parece que el responsable sería un cambio regional de la circulación alrededor de Groenlandia, que permite la intrusión de agua más caliente favoreciendo este proceso. No está claro si estos cambios de circulación están relacionados con la variabilidad interdecadal de la Oscilación del Atlántico Norte (NAO por sus siglas en inglés) o una tendencia asociada con el calentamiento asociado a los gases de efecto invernadero. Las proyecciones de los modelos a lo largo de este siglo indican que el descuadre entre acumulación y pérdida podría continuar a lo largo del siglo XXI. Hay que notar, sin embargo, que la recongelación del agua deshelada en las capas más altas de la capa de hielo es un mecanismo que podría reducir la pérdida de masa, mientras que por el contrario un incremento del deshielo superficial podría facilitar la lubricación de la base del hielo favoreciendo un aumento de la pérdida de masa. Así, las proyecciones sobre el aumento del nivel del mar por la contribución del hielo polar varían entre los 2 cm y 23 cm a finales de siglo.

La dinámica de estos mecanismos no parece estar afectada por transiciones bruscas que hicieran pensar en un colapso de la masa de hielo de Groenlandia que diera lugar a cambios importantes y bruscos del nivel del mar. Sin embargo, los modelos conceptuales parecen indicar la existencia de un barrera para la temperatura que si se sobrepasara durante suficiente tiempo (miles de años) la pérdida de hielo sería irreversible en estas escalas de tiempo incluso aunque el clima reestableciera su estado entretanto. En ese caso la contribución sería equivalente a un aumento del nivel del mar de 7 m. Intentos de modelar el Plioceno (5-6 millones de años a.c.), cuando se daban temperaturas y niveles de CO_2 similares a los actuales, indican que temperaturas de 2-3.5ºC superiores a los niveles pre-industriales producen una desaparición casi completa de la capa de hielo (IPCC, 2013, a partir del trabajo de Hill et al. 2010). Sin embargo estos valores deben tomarse con precaución dado que la situación climática del Plioceno a escala regional no es comparable según un estudio reciente de los propios autores (Hill, 2014).

En el caso del hielo Antártico la situación es diferente. Al igual que en Groenlandia la cantidad de masa de hielo responde a un balance entre la acumulación de nieve y el flujo de hielo hacia zonas más periféricas en contacto con el mar circundante en escalas milenarias. A diferencia de Groenlandia, la Antártida tiene temperaturas superficiales más frías que no inciden en el deshielo superficial y las pérdidas se producen mayormente por desprendimiento de icebergs sobre el mar (ice calving) y por mezcla submarina.

Como ya hemos avanzado anteriormente en varias zonas de la Antártida la capa de hielo se prolonga de forma continua como una plataforma por encima del mar. En esas áreas, esporádicamente la cobertura de hielo puede romperse liberando placas enormes que se separan de la masa de hielo principal como ocurrió en 2002 con la plataforma Larsen B. Por debajo de esta placa, el hielo se derrite en contacto con el agua de mar. En esas áreas la placa de hielo se sitúa sobre un fondo de roca que puede estar sumergido hasta 2000 y 3000 m de profundidad, bastante por debajo de la profundidad del mar adyacente (véase el esquema de la figura 2).

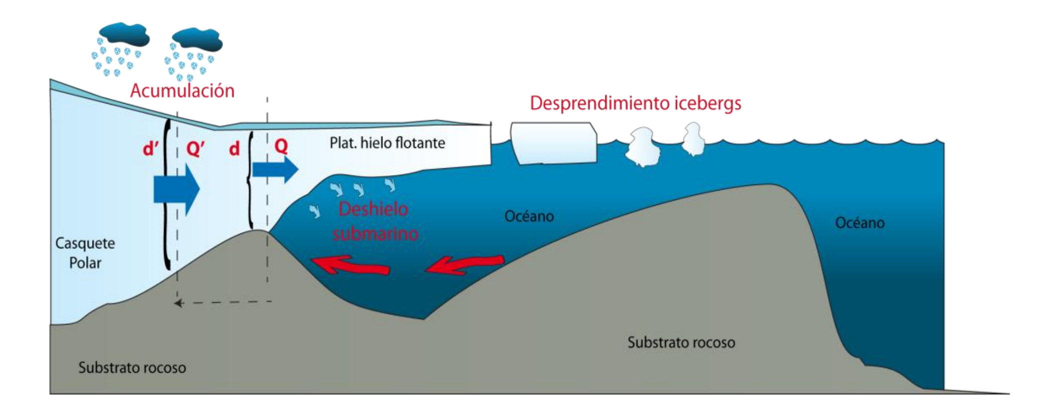


Figura 2: *Esquema sintético de la Inestabilidad de la Capa de Hielo Marina (MIS). En el diagrama d_e y Q_e son los valores del flujo de hielo y grosor en la posición de equilibrio. Si este punto retrocediera un poco por un aumento de la temperatura de las aguas submarinas por debajo de la plataforma de hielo flotante, la nueva posición de equilibrio sería tal que Q'> Q dado que d'>d en ciertas regiones de la Antártida.*

A mediados de los 70, se propuso que el balance del flujo de hielo dependía de la posición de la línea donde empezaba la plataforma de hielo flotante, es decir el límite máximo donde el agua marina estaba en contacto con el hielo continental. Si en ese punto el grosor de la capa de hielo aumentaba, el flujo de hielo aumentaba. Análisis recientes más rigurosos a partir de las leyes mecánicas que rigen la dinámica del flujo de las capas de hielo han podido corroborar este hecho y determinar, bajo ciertas aproximaciones, la ley de escala entre el grosor de hielo y el flujo en ese punto (existe una discusión más detallada en IPCC, 2013). Si dicha línea se retira a una posición de mayor profundidad del substrato rocoso, entonces el flujo de salida debe aumentar, dado que en esa nueva posición la capa de hielo resulta ser más gruesa, y así sucesivamente en lo que se denomina Hipótesis de Inestabilidad de la Capa de Hielo Marina (MIS acrónimo en inglés). Este proceso podría verse desencadenado por un mayor deshielo asociado al calentamiento de las aguas justo debajo de la plataforma de hielo o también por la aparición en la superficie de puntos de fusión que indujeran fracturas verticales de la masa de hielo separando la plataforma. Medidas tras el colapso de la plataforma Larsen B han corroborado ésta sutil conexión entre la plataforma de hielo y su conexión con la masa continental.

En definitiva la evolución de la capa de hielo Antártica es compleja y las variaciones del balance entre acumulación de nieve y flujo saliente, dependen a su vez de las condiciones de la circulación atmosférica y oceánica circundantes así como de las diferentes características de las diferentes áreas geográficas de la Antártida. En cuanto a los registros de medidas disponibles, los datos a partir de radar desde satélite indican que la caída de nieve ha aumentado en la región más hacia el Este (aunque medidas derivadas del cambio de la gravedad no señalan un aumento significativo), que ha habido un aumento de flujo saliente en áreas localizadas y un aumento bien documentado de colapso de la capa en la región al Norte. Así el balance neto entre las diferentes regiones de la Antártida indica una contribución neta del ritmo de variación del nivel del mar de 0.27 mm/año durante el período 1993-2010. Sin embargo, las proyecciones previstas para el siglo XXI, indican una contribución negativa en tanto en cuanto se observa un aumento significativo de nieve acumulada. La explicación sería que un calentamiento atmosférico favorecería en esta región un aumento del aire

húmedo induciendo un aumento notable de la acumulación de nieve. Estas proyecciones deben tomarse con cautela en vista de que, a día de hoy, los intrincados procesos asociados a la MIS son difíciles de ser integrados en los modelos de clima.

Pero volvamos a las cuestiones que han motivado este breve y muy somero repaso acerca de lo que sabemos de la evolución de los casquetes polares. Las consecuencias más evidentes del deshielo de los casquetes es sin duda su contribución a la subida del nivel del mar medio. Sin embargo, no parecen haber razones que dieran a pensar en un cambio brusco del mismo al menos proveniente del deshielo de Groenlandia pero quizás sí en el caso de la Antártida asociado a la MIS. Además debemos ser conscientes que el nivel del mar varía de forma regional e incluso local por otros factores por lo que el impacto sería muy diferente según las diferentes regiones del planeta. Recordemos que hoy por hoy la mayor contribución al nivel del mar es el efecto termostérico (dilatación térmica) por el calentamiento de la masa oceánica. Frente a un cambio brusco, que no nos atrevemos a cuantificar dadas las incertidumbres que tenemos tanto en término de medidas como de conocimiento preciso de los procesos involucrados, probablemente el impacto sobrepasaría sobradamente el efecto termostérico. Además, el deshielo en general también produce una retroalimentación positiva en relación al albedo, es decir esa porción de la radiación solar que se refleja hacia el espacio en las superficies con hielo y nieve, lo cual incidiría en una aceleración del calentamiento global. Un tercer efecto del deshielo tendría que ver con modificaciones de la circulación general del océano afectada por un aporte importante de agua dulce. En el caso de Groenlandia entre sus posibles efectos estaría la modificación de las condiciones de formación de agua densa en el Atlántico Norte, una de las componentes esenciales del ciclo natural del océano para disipar el exceso de calor del Ecuador. Sin embargo, como hemos visto los datos y el conocimiento que disponemos por ahora indican que un deshielo brusco masivo en Groenlandia no es una cuestión que se prevea pueda ocurrir a corto plazo.

Ya para acabar queda por ver qué parte de este deshielo es achacable a la acción humana. La respuesta es difícil de contestar cuantitativamente. Cómo hemos visto en los procesos anteriores no solo interviene una acción directa del balance radiativo en forma de aumento de temperatura sino que hay implicaciones de la variabilidad de la circulación atmosférica y oceánica regional. La acción humana se traduce en una mayor concentración de gases de efecto invernadero que contribuyen en gran medida al calentamiento y por tanto pueden influir en términos de un calentamiento de las aguas y aire circundantes de estas regiones. Sin embargo hemos visto que los procesos de deshielo tienen mucho que ver con el ciclo del agua que aunque bien conocido cualitativamente puede modificar en gran medida los balances de acumulación y flujo saliente de hielo en ambos hemisferios de manera no equivalente. Aunque hoy en día se puede asegurar que el calentamiento en el hemisferio Norte favorece el deshielo superficial tanto del Ártico como de la capa sobre Groenlandia, dicho calentamiento no parece ser tan relevante en el hemisferio Sur. Allí el deshielo neto observado parece venir más por un progresivo calentamiento de las aguas del hemisferio sur y cambios de la circulación atmosférica que, al menos en los estudios revisados, no permiten discernir en qué medida la acción humana contribuye o ha contribuido.

Referencias:
Hill D.J., 2015: The non-analogue nature of Pliocene temperature gradients. Earth and Planetary Science Letters, 425, 1, 232–241
Hill D.J., Dolan A.M. Dolan, Haywood A.M., Hunter S.J. and Stoll D.K, 2010: Sensitivity of the Greenland Ice Sheet to Pliocene sea surface temperatures. Stratigraphy, vol. 7, nos. 2-3, pp. 111-122,
Jevrejeva S., Moore J.C., Grinsteda A., Matthews A.P. and Spada G., 2014: Trends and acceleration in global and regional sea levels since 1807. Global and Planetary Change, 113, 11-22.
IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental. Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Emilio García Ladona **Joaquim Ballabrera Poy**
Doctores en Ciencias Físicas
Dept. de Oceanografía Física y Tecnológica
Instituto de Ciencias del Mar (ICM-CSIC)

Capítulo 30
¿De dónde vino el aire que respiramos?

(Por Raghuveer Parthasarathy)

Cada vez que respiramos, usamos moléculas producidas por otros innumerables seres vivos, y también por procesos geológicos que se remontan hace miles de millones de años. Desde la perspectiva de las células de nuestros cuerpos, la molécula más importante en el aire es el oxígeno. Cada una de nuestras células usa el oxígeno para realizar reacciones químicas que liberan energía de los azúcares, grasas, y proteínas que consumimos, soltando dióxido de carbono y agua como productos. Sin el consumo de oxígeno, seríamos incapaces de extraer la energía de la comida. Por lo tanto, preguntaremos primero ¿de dónde vino el oxígeno de la atmósfera de la Tierra?, y a continuación examinaremos otros componentes del aire que respiramos.

A primera vista, poco tienen que ver la vida de las plantas con el proceso respiratorio. Sin embargo, los dos están marcadamente ligados. Las plantas absorben suficiente energía por un proceso llamado fotosíntesis, la conversión de energía contenida en la luz en energía almacenada en enlaces químicos. Mientras el dióxido de carbono y el agua se consumen, el oxígeno y las moléculas orgánicas ricas en energía se producen (efectivamente, es el revés del proceso respiratorio del que se sirven nuestras células al consumir el oxígeno). La actividad de los organismos realizando la fotosíntesis es la fuente del oxígeno en la atmósfera, el oxígeno que mantiene vivos a los animales, incluidos nosotros. La mayoría del oxígeno atmosférico viene de los océanos, producido por organismos microscópicos llamados fitoplancton. El resto viene de plantas terrestres. Es interesante pensar en el rango de tamaños de las criaturas que nos brindan el oxígeno, desde la bacteria fotosintética que mide la millonésima parte de un metro hasta los árboles más grandes que miden más de cien metros de altura.

El oxígeno es bastante abundante, componiendo una quinta parte de la atmósfera del planeta. Sin embargo, no ha sido siempre así. La tierra tiene usos 4.5 mil millones de años, y los organismos unicelulares capaces de la fotosíntesis surgieron después de uno o dos mil millones de años. Aun después, el oxígeno era muy escaso en el aire. El oxígeno es un elemento muy reactivo, y durante cientos de millones de años, moléculas de oxígeno rápidamente se absorbían por minerales, creando compuestos como óxido de hierro (moho) que observamos aun hoy en formaciones rocosas antiguas. Hace unos 2.4 mil millones de años, la tierra pasó por un periodo al que se refieren los geólogos como el Gran Evento Oxigenizante. Aquí, las medidas geológicas, muestran un gran aumento constante en el oxígeno atmosférico. Las causas, sea un aumento en la producción de poblaciones de organismos fotosintéticos o bien el consumo reducido del oxígeno por procesos geoquímicos, están todavía oscuras. Independientemente de su causa, el resultado fue un incremento en la abundancia del oxígeno libre, hasta niveles todavía inferiores a un décimo de los niveles actuales. Hace unos, entre 1000-500, millones de años la tierra pasó por un segundo evento oxigenizante, con niveles de oxígeno que iban variando entre 15 y 30% a lo largo de los últimos cientos de millones de años. Estos

niveles del oxígeno hicieron posible el desarrollo de organismos multicelulares, tales como los animales.

¿Y los otros componentes del aire que respiramos? El nitrógeno compone el 78% de la atmósfera. A todos los organismos les hace falta, es un componente de proteínas, ADN, y otras moléculas biológicas. Sin embargo, no podemos ingresar directamente el nitrógeno del aire, siendo un gas inerte. Ciertas bacterias, por ejemplo especies que viven en el suelo y forman relaciones simbióticas con las plantas, poseen la capacidad de 'fijar' el nitrógeno del aire, combinándolo con el hidrógeno para formar el amonio, que luego puede convertirse en otros compuestos útiles. Hace unos cien años, los químicos Fritz Haber y Carl Bosch inventaron una manera de fijar artificialmente el nitrógeno, llevando al desarrollo de abonos artificiales y un vasto incremento de la productividad agrícola, y por ende, un vasto incremento de la población humana en la Tierra. La descomposición de sustancia orgánica, por varios procesos, emite el nitrógeno a la atmósfera. Este 'ciclo del nitrógeno' determina el fluir de este importante componente atmosférico. El oxígeno y el nitrógeno son los componentes más abundantes de la atmósfera. Hay también otros constituyentes, que pueden tener un gran impacto en la Tierra incluso en cantidades pequeñas. Uno particularmente importante es el dióxido de carbono, que ya hemos encontrado como producto de la respiración celular. El dióxido de carbono es buen absorbente de la radiación infrarroja, una forma invisible de luz emitida por objetos cálidos, como la superficie de la Tierra. La absorción de la luz infrarroja por el dióxido de carbono hace que la superficie de la Tierra sea más cálida de lo que debiera, un fenómeno conocido como el Efecto Invernadero. (El vapor de agua es otro gas que también contribuye al Efecto Invernadero). De manera muy semejante a la respiración celular, quemar las moléculas orgánicas como los combustibles fósiles (petróleo, carbón, y gas natural) suelta dióxido de carbono en la atmósfera. La cantidad de combustibles fósiles que quemamos es tan enorme que tiene un impacto significativo en la concentración atmosférica del dióxido de carbono, que ha incrementado de unas 280 partes por millón a unas 400 partes por millón desde los albores de la Revolución Industrial, hace unos 200 años. Esta cantidad tan grande, y sigue en aumento, conduce a un Efecto Invernadero aun más fuerte, y por lo tanto un calentamiento global que ya comienza a tener consecuencias graves para el clima del planeta.

Pensar en el dióxido de carbono, sin embargo, nos lleva a pensar tanto en el pasado lejano como en el futuro inminente. Mucho antes del origen de la vida en la Tierra, la atmósfera era radicalmente diferente de lo que es hoy. Hidrógeno y helio, los elementos más abundantes en el universo, formaron la primera atmósfera del planeta, cuando el planeta se formó de nubes de polvo y gas. Sin embargo, estos elementos ligeros fácilmente se escaparon de la atracción gravitatoria del planeta. La actividad volcánica en la Tierra temprana soltó vapor de agua, amonio, y dióxido de carbono. El vapor de agua se condensó para formar océanos. Parte del dióxido de carbono se disolvió en estos océanos —experimentamos a menudo esta solubilidad del dióxido de carbono en bebidas carbonizadas— donde era recurso para los organismos fotosintéticos y paulatinamente llevó a la abundancia del oxígeno atmosférico, así cerrando el círculo de nuestra historia.

Gracias al dr. **David Wacks** (Professor,University of Oregón) por una traducción precisa de este capítulo.

Raghuveer Parthasarathy
Doctor en Física
Profesor Titular de Física, University of Oregón (EEUU)

Capítulo 31
¿Qué es el efecto invernadero?

(Por Emilio García Ladona y Joaquim Ballabrera Poy)

El Sol irradia energía electromagnética que llega a la superficie tras interaccionar con los gases que componen la atmósfera de la Tierra. Para entender qué ocurre y cómo se produce ésta interacción debemos saber cuáles son las propiedades de absorción y emisión tanto de la fuente emisora como de los elementos que reciben la energía solar. En primer lugar debemos tener una idea de qué gases conforman la atmósfera. En la tabla 1 hemos listado los componentes principales de la atmósfera terrestre. Se observa que el N_2 y el O_2 son los más abundantes seguidos, en menor medida, por una serie de componentes con una fracción de volumen menor del 1%, excepto el vapor de agua que puede variar hasta un 5% según las condiciones atmosféricas.

La radiación solar que llega a la parte superior de la atmósfera se puede aproximar por la radiación de un cuerpo negro a una temperatura 5525 K (figura 1) y centrada en la banda de luz visible, pero conteniendo una fracción importante de ultravioleta e infrarrojo corto (< 3 micras). Tras atravesar la atmósfera, la radiación incidente se ve atenuada en un 70-75 % en la superficie terrestre y la mayor atenuación se produce en la banda de ultravioleta y visible. Esto es debido esencialmente a la absorción del O_3 en la estratosfera principalmente así como una cierta dispersión de la radiación (dispersión de Rayleigh). En la banda inferior del infrarrojo (~0.6-2 micras) la curva muestra unos lapsus o pozos que se corresponden esencialmente con las diferentes bandas de absorción del vapor de agua. Así, el vapor de agua actúa como un primer filtro radiativo absorbiendo parte de la radiación entrante.

Tras atravesar la atmósfera, la radiación restante interacciona con la superficie de la Tierra siendo reflejada o absorbida. La mayor parte de esta radiación es absorbida por el océano. Allí la penetración disminuye fuertemente y prácticamente en el primer metro solamente queda radiación en la banda visible alrededor de 462 nm (~ 0.4 micras) típicamente el color azul de aguas transparentes, y a 100 m de profundidad la atenuación es prácticamente total. A pesar de esta fuerte atenuación entre 0-100 m, la radiación solar es suficiente para sustentar el desarrollo de la fotosíntesis de prácticamente toda la vida vegetal marina (fitoplancton).

La absorción de la radiación electromagnética en el mar es una fuente de energía que calienta las capas superficiales y contribuye a potenciar las corrientes oceánicas a gran escala con el fin de redistribuir y homogeneizar esta energía por todo el planeta. Al hacer un promedio de la temperatura de la superficie de la Tierra nos encontramos con un valor característico de unos 15-17 ºC que, a su vez, por la ley de Planck irradia energía electromagnética. El espectro de emisión (curvas suaves en negro, azul y lila en la figura 1) está en la banda del infrarrojo "largo" (~3-60 micras). Sin embargo vemos que el espectro de energía que escapa al espacio exterior tras haber atravesado la

atmósfera tiene una forma muy alejada de las curvas de emisión teóricas. Esto sucede porque el tipo de radiación emitida coincide con varias bandas de absorción características de muchos más componentes atmosféricos (CO_2, CH_4,…) dejando libre tan solo esa ventana en el rango 8-15 micras ("*ventana atmosférica*") para la radiación saliente (esa es la banda que utilizan la gran mayoría de los satélites de observación de la Tierra para escudriñar la superficie terrestre). Cómo podemos observar básicamente los gases que absorben en esas bandas son el vapor de agua nuevamente, el CO_2, el CH_4 y el óxido nitroso NO_2.

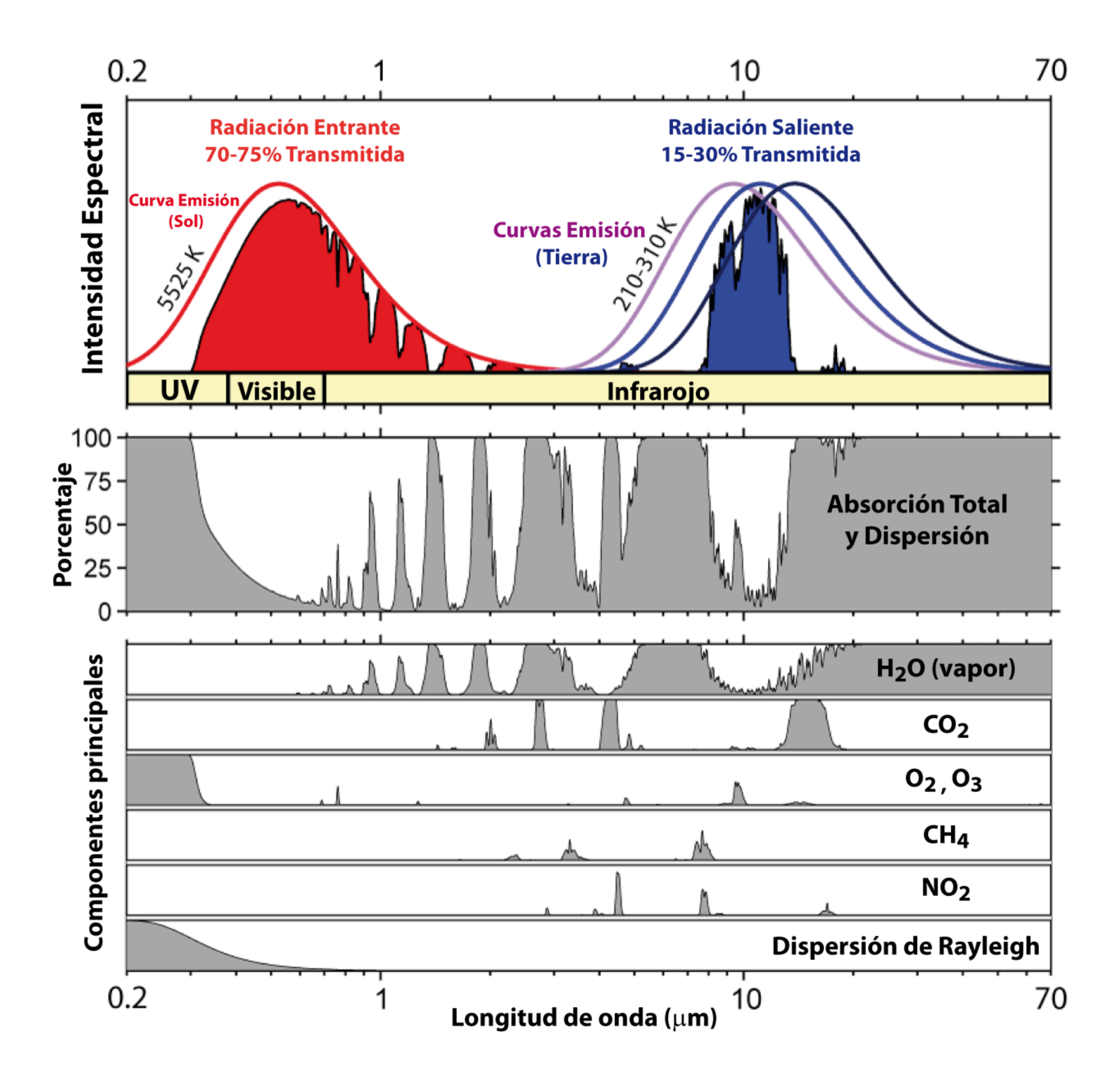


Figura 1: *Intensidad espectral de la radiación de entrada y salida de la Tierra y curvas de absorción de radiación total y de cada uno de los componentes atmosféricos principales que intervienen en el balance radiativo de la Tierra (incluida la dispersión de radiación por las moléculas, dispersión de Rayleigh). Las curvas continuas corresponden a las curvas de emisión de la Ley de Planck para la temperatura característica del Sol (5525 K) y 3 valores de la superficie de la Tierra entre -63ºC (210 K) y 37ºC (310 K). Datos basados en la calculadora espectral (http://www.spectralcalc.com) a partir de la base de datos espectroscópicos HITRANS2004. © Robert A. Rhode reproducida aquí bajo licencia GNU versión 1.2 de la Free Software Foundation.*

Por tanto estos gases absorben parte de esta radiación saliente y la reemiten, una parte de retorno a la superficie y otra hacia capas superiores de la atmósfera, y así hasta que abandona finalmente la atmósfera. Es la energía de la radiación reemitida por estos gases añadida a la de la radiación entrante que contribuye a establecer la temperatura característica del planeta y que se denomina comúnmente como *efecto invernadero*. Llegado a este punto debemos precisar acerca de este concepto que a nuestro modo de ver podemos considerar inadecuado. Una experiencia cotidiana muy común al entrar en un coche estacionado que haya estado expuesto al Sol durante varias horas, con puertas y ventanas cerradas, es que la temperatura del interior es muy alta y que el tablero de mandos y el volante literalmente queman al tocarlos con la mano. La radiación solar penetra dentro del habitáculo calentando todos los objetos y superficies además del aire en su interior. Esta sencilla observación es análoga a la que se produce en locales y verandas exteriores encerradas con cristaleras y es bien conocido de los invernaderos para proteger plantas que pueden ser sensibles a temperaturas ambientes excesivamente frías. Sin embargo, y volviendo al coche, si dejamos abierta un poco una de las ventanas, la temperatura dentro del habitáculo será menor que si estuviera cerrada. La causa es que la ventana abierta facilita el intercambio de masas de aire y por ende de temperatura entre el interior y exterior vía la convección, que es una forma efectiva y rápida de intercambiar calor. Es decir, básicamente el mayor calentamiento del interior de los habitáculos tipo invernaderos no se debe tanto a un atrapamiento de la radiación y reemisión de la radiación electromagnética por las paredes que lo delimitan, sino por la supresión del intercambio convectivo con el aire exterior.

En el caso de la atmósfera terrestre el calentamiento no se produce por una supresión del intercambio convectivo, dado que es un sistema abierto, sino básicamente por la absorción y reemisión de energía de los gases atmosféricos hacia las capas más bajas de nuevo incidiendo en el balance neto entre la radiación entrante y la radiación saliente. Según sea este balance, la temperatura terrestre aumentará o disminuirá o permanecerá constante en caso de tener un perfecto balance radiativo. Cuando este balance se rompe la temperatura varía hasta alcanzar un nuevo estado de equilibrio radiativo con modificaciones que afectan de una manera u otra al sistema climático. De hecho la historia de la Tierra ha sido un continuo cambio de patrones climáticos a veces fruto de eventos catastróficos (grandes explosiones volcánicas, impacto de asteroides, etc.) que han modificado el balance radiativo a escalas interanuales, decadales, seculares, milenarias....

Evidentemente el desequilibrio del balance radiativo puede variar o bien porque la radiación entrante varía (relacionado con la propia dinámica de Sol) o bien porque la radiación saliente varía (alteración de los componentes y condiciones que más contribuyen a la emisión). Como apuntábamos antes, la historia de la Tierra tal y como la conocemos, nos ha enseñado que ambas componentes han ido fluctuando, configurando un sistema climático variable dando lugar a periodos fríos y cálidos, a veces extremadamente fríos (conocidos como superglaciaciones o Tierra Bola de Nieve del inglés *SnowballEarth*). No es de extrañar que una parte importante de estos cambios estén asociados al ciclo del agua, dado que podemos observar cómo el vapor de agua es uno de los gases que mayor influencian el balance radiativo al absorber simultáneamente radiación entrante y saliente. La prueba más evidente de cómo el vapor de agua afecta de forma apreciable al balance radiativo, la podemos constatar en nosotros mismos cuando observamos las variaciones de la temperatura entre los días nublados y los días despejados, sobre todo durante la noche. En noches cubiertas la temperatura en la superficie es por lo general más cálida ya que la absorción y

reemisión de la energía saliente por las nubes atenúa el enfriamiento que se produciría si los cielos estuvieran despejados. Resulta cuando menos curioso, rayando la falta de rigurosidad, que el papel del vapor de agua sea uno de los grandes ausentes de los debates o presentaciones que tan a menudo se oyen en medios de comunicación o conferencias divulgativas acerca de las causas y consecuencias del cambio climático en la era moderna.

Ciertamente, la acción humana ha incidido en el sistema climático, entre otros aspectos, contribuyendo artificialmente a aumentar la concentración de los gases que intervienen en el balance radiativo. Uno de los efectos más evidentes de estos gases es la realimentación positiva precisamente a través del vapor de agua. Así las bandas de absorción del CO_2, en particular algunas que no se superponen con las del vapor de agua (alrededor de las 4 micras) son una aportación neta al balance radiativo. Inducen un aumento de la temperatura, favoreciendo a su vez un aumento de la humedad y por tanto potenciando el papel del vapor de H_2O. El papel que juega el vapor de H_2O en la atmósfera contribuyendo con retroalimentaciones positivas y negativas en el balance radiativo es uno de los aspectos esenciales para entender plenamente de qué manera ha evolucionado, evoluciona y evolucionará el clima en el futuro.

En el resumen del capítulo 9 del último informe del IPCC panel (IPCC, 2013) se indica que los modelos climáticos reproducen razonablemente bien las tendencias genéricas de la evolución del clima terrestre durante el pasado siglo (mayormente utilizando la evolución de la temperatura media en superficie). Sin embargo, también señalan que hay discrepancias con la evolución del clima de los últimos 30 años donde la temperatura no ha aumentado como habían predicho los modelos en el informe anterior, pese al continuo aumento de emisión de gases invernadero. Dejando aparte la representatividad de los modelos actuales de clima, la hipótesis que parece tomar mayor relevancia es una mayor acumulación de calor en el océano.

¿Por qué entonces la temperatura de la Tierra no ha aumentado en proporción a la acumulación de calor y el aumento continuado de los gases de efecto invernadero, en particular del CO_2? La explicación radica en que el agua tiene una capacidad calorífica mucho mayor que el aire, unas 4000 veces más grande. Esto significa que si tenemos dos volúmenes iguales de aire y agua y queremos elevar 1ºC la temperatura de ambos, necesitamos unas 4000 veces más de energía calorífica para el agua que para el aire. Lo mismo sucede entre el agua y las superficies rocosas en general aunque las diferencias son menores. Así ese aumento de calor en el océano no ha supuesto un aumento considerable de la temperatura en tanto en cuanto el océano además constituye la mayor superficie del planeta. En esto radica el papel del océano como regulador del clima que permite tener en zonas costeras variaciones de temperatura muy suaves respecto a las grandes variaciones de temperatura características de las zonas interiores de los continentes. En cierto modo la evolución futura del clima, como ha sido sin duda en el pasado, va a depender en gran medida de cómo responda y evolucione el océano ante esta mayor acumulación de calor.

Constituyente	**Porcentaje en volumen**
Nitrógeno(N_2)	78.08 %
Oxígeno (O_2)	20.95 %
Argón (Ar)	0.93 %
Vapor de Agua (H_2O)	0-5 %
Dióxido de Carbono (CO_2)	380 ppm
Neón (Ne)	18 ppm
Helio (He)	5 ppm
Metano (CH_4)	1.75 ppm
Kriptón (Kr)	1 ppm
Hidrógeno (H_2)	0.5 ppm
Óxido Nitroso (NO_2)	0.3 ppm
Ozono (O_3)	0-0.1 ppm

Tabla 1*: Principales constituyentes de la atmósfera ordenados según el porcentaje en volumen (ppm, partes por millón, 10.000 ppm = 1%). En negrita los constituyentes más relevantes para el balance radiativo de la Tierra (gases de efecto invernadero). Fuente: J. Wallace, P. Hobbs, AtmosphericScienceanIntroductorySurvey*

Emilio García Ladona **Joaquim Ballabrera Poy**
Doctores en Ciencias Físicas
Dept. de Oceanografía Física y Tecnológica
Instituto de Ciencias del Mar (ICM-CSIC)

Capítulo 32
¿Cuál será la causa más probable de la extinción humana?
(Por Felipe Zapata Ruiz)

En la historia de nuestro planeta han ocurrido cinco grandes extinciones masivas, que se caracterizan por un decrecimiento acelerado de la variedad de organismos presentes en nuestro planeta. Algunos de estos fenómenos se han atribuido a cambios climáticos extremos, actividad volcánica, supernovas o impacto de asteroides, entre otros. Sin embargo estas no son las únicas extinciones que han ocurrido y ocurren en nuestro planeta, por ejemplo, la extinción masiva del Holoceno empezó aproximadamente 10000 antes de cristo y continúa hasta la actualidad donde miles de especies animales han desaparecido, principalmente por el cambio climático y la acción del hombre. Tomando como punto de partida las extinciones masivas de las especies y la extinción del Holoceno, vamos a clasificar las causas más probables de la extinción humana en dos grandes grupos: causas relacionadas con la acción del hombre o antropogénicas y causas naturales.

32.1 Causas Antropogénicas

En esta categoría apocalíptica se encuentran los siguientes tópicos que causan revuelo entre científicos, gobiernos y el público en general, quienes debaten acaloradamente sobre el verdadero alcance de los siguientes fenómenos: calentamiento global, resistencia a los antibióticos y la muerte de las abejas. Esta lista ignora de manera flagrante los desastres nucleares o conflictos entre países, que podrían diezmar la población humana pero que están lejos de una posible extinción (al menos así lo espero yo).

El calentamiento global es uno de los tópicos más ardientes en todos los ámbitos de la sociedad y para evitar discusiones innecesarias empecemos por recalcar que este fenómeno es causado por los humanos como ha sido establecido por la panel intergubernamental para el cambio climático (IPCC) por sus siglas en ingles, donde se ha declarado que: “ Existe más de un 90% de certeza que la emisión de los gases de invernadero resultantes de las actividades humanas han causado la mayor parte del aumento en las temperaturas observadas desde la mitad del siglo XX”. El cambio climático no solo involucra un aumento en las temperaturas medias del planeta generando sequías y hambrunas (por ejemplo el verano del 2016 es el mas cálido desde que empezamos los registros de la temperatura a mediados del siglo XIX), sino que dispara un conjunto de cambios a escala global que repercute en la naturaleza. Al aumentar la temperatura, se provoca un deshielo masivo de los casquetes polares y los glaciares que aumentan el nivel del mar y pone en peligro de inundación las ciudades costeras, además de verter al mar millones de litro de agua potable, reduciendo su salinidad y afectando a los organismos vivos de agua salina. También resulta que debajo

de las capas de hielo en el hemisferio norte hay atrapados miles de toneladas de metano en estado gaseoso, que al derretirse las capas de hielo escapan al aire, y vosotros os preguntareis que tiene que ver el metano con todo esto, pues resulta que el metano es un gas de invernadero mucho peor que el CO_2, causando una aceleración en el calentamiento. Recientemente una ola de calor en Siberia ha descongelado una cepa bacteriana de ántrax que ha provocado la muerte a 40 indígenas de la zona y cientos de ciervos.

Además sabemos que la solubilidad del oxigeno en agua es inversamente proporcional a la temperatura, es decir que a medida que los océanos se calientan su nivel de oxígeno disminuye, poniendo en riesgo los corales marinos de lo que dependen millones de peces y nosotros, por supuesto. Como si esto fuera poco, las corrientes de aire en nuestro planeta se comportan como un sistema cerrado, aproximadamente como una olla a presión y a mayor temperatura mayor presión, lo que hace que los huracanes, tormentas y demás fenómenos que resultan de la diferencia de presión y temperatura de las corrientes de aire sean cada vez más violentos. Como dato curioso, uno de los mayores productores de metano en nuestro planeta es el estómago de las vacas que libera grandes cantidades de este gas durante la digestión, así que la próxima vez que estés comiéndote una deliciosa hamburguesa en un día caluroso ¡recuerda la venganza de las vacas!

Nuestra segunda causa de una posible extinción humana es la resistencia a los antibióticos, título que recuerda a una película de ciencia ficción de bajo presupuesto, pero la realidad ha mostrado que la resistencia a los antibióticos es un problema muy complejo. Gracias a Louis Pasteur y su investigación en la segunda mitad del siglo XIX, dimos pasos agigantados en el entendimiento de los microorganismo y su relación con las enfermedades, seguidamente Alexander Fleming cambió para siempre la historia del a humanidad cuando introdujo la penicilina como método efectivo para combatir infecciones bacterianas. Desde entonces hemos venido utilizando de forma rutinaria diferentes antibióticos para combatir enfermedades que han matado a millones de personas en la historia de la humanidad. El problema con los antibióticos empieza cuando intentamos matar un mosquito a cañonazos. Cuando tenemos una gripe o catarro en lugar de dejar que nuestro sistema defensivo se encargue naturalmente de ello, tomamos antibióticos para aliviar prontamente el malestar y una vez nos sentimos mejor dejamos los antibióticos de lado y seguimos con nuestras vidas, dejando con vida a parte de los microorganismos que causaban el catarro. Ingenuamente creíamos haber eliminado el causante del problema, pero parte de los microorganismos agresores han sobrevivido y peor aun ahora saben cómo resistir al antibiótico. Los microorganismos se reproducen muy rápido (algo así como un centenar de generaciones por día) y las nuevas generaciones tienen pequeñas mutaciones que los diferencias de sus progenitores, estas mutaciones pueden ser, por ejemplo, la resistencia a ciertos antibióticos. Como resultado se ha creado una generación de microorganismos resistentes a una familia de antibióticos, si una persona cae enferma con unos de estos microorganismos resistentes debe aplicarse un antibiótico más fuerte y el ciclo de la resistencia puede repetirse.

El mal uso de los antibióticos lleva dándose desde mediados del siglo XX en países desarrollados dado que los antibióticos están al alcance de nuestra mano. El resultado es entonces que hemos producido un conjunto de microorganismos que son resistentes básicamente a todos los antibióticos que conocemos y que de no tomarse medidas prontamente podríamos llegar a morir de enfermedades comunes que años

atrás parecían haberse extinguido de la tierra. Muchos de vosotros os preguntaréis si no es posible crear antibióticos más potentes y la respuesta no es tan sencilla puesto que la mayoría de los antibióticos actuales son moléculas pequeñas que presentan interacciones relativamente simples con los microorganismos, son fáciles de producir en masa y testear. Mientras los hipotéticos futuros antibióticos más potentes involucrarían una mayor complejidad tanto en la interacción con los patógenos como en su producción en masa. La próxima vez que pienses en usar algún antibiótico para curar un catarro, piénsalo dos veces y el resto de la humanidad te lo agradecerá.

La tercera posible causa de extinción debida a posibles razones humanas es la muerte acelerada de millones de abejas que hemos venido observando en los últimos años. La desaparición de millones de abejas se conoce como el "colapso desordenado de la colonia", cuya causa está aún en debate pero se asocia a varios factores como la utilización de pesticidas, el *virus de las alas deformes,* el cambio climático entre otras. Aparte de producir la deliciosa miel, las abejas polinizan la mayoría de las plantas que consumimos los humanos o que utilizamos como alimento para nuestro ganado, por lo tanto la desaparición repentina de millones de abejas significa la subsecuente muerte de millones de plantas y seguidamente de millones de humanos. Un comentario atribuido a Albert Einstein dice que: "si las abejas desaparecieran de la superficie del planeta al hombre solo le quedarían cuatro años de vida".

32.2 Causas Naturales

Dentro de las causas naturales que podrían poner en peligro nuestra existencia como especie se encuentran los siguientes eventos hipotéticos: erupciones volcánicas a gran escala, colisiones con objetos pequeños del sistema solar o un estallido de radiación gamma.

Erupciones volcánicas de dimensiones épicas han sido asociadas con la extinción masiva del Triásico-Jurásico. Estas erupciones pudieron haber liberado grandes cantidades de gases invernadero y cenizas que pudieron generar un calentamiento global acelerado o las cenizas pudieron haber ocultado el Sol durante meses evitando que las plantas llevarán a cabo la fotosíntesis por lo que los animales hubieran perecido por falta de alimentos.

Aunque nunca ha sido reportado ninguna colisión con un cuerpo celeste que tuviese consecuencias graves para la humanidad, no quiere decir que un evento de este tipo no pueda suceder y traiga consigo gran devastación entre la población humana y sus campos de cultivo, sino también puede levantar una gran cantidad de partículas que pueden oscurecer el cielo por un tiempo significativo destruyendo plantas y generar una hambruna.

Por otra parte un estallido de radiación gamma o rayos x, provenientes de la explosión de un estrella súper masiva, podría destruir completamente la capa de ozono, dejando la superficie de la tierra expuesta a los peligrosos rayos ultravioleta provenientes del Sol, resultando en una catástrofe para toda la vida en el planeta.

Bibliografia:
IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
http://news.nationalgeographic.com/news/2013/03/130321-triassic-mass-extinction-volcano-paleontology-science/
http://www.iflscience.com/health-and-medicine/zombie-anthrax-outbreak-in-siberia-blamed-on-thawedout-infected-reindeer-corpse/

Felipe Zapata Ruiz
Doctor en Química
Investigador Postdoctoral, Vrije Universiteit, Amsterdam

Capítulo 33
¿Cuál es la unidad de tiempo más pequeña posible?
(Por José Edelstein)

Todas las culturas humanas desarrollaron sistemas para medir el paso del tiempo, ese flujo que experimentamos en la sucesión de eventos, que los conecta causalmente y que, según parece, no tiene vuelta atrás. Cualquier acontecimiento periódico, que se repitiera con un ritmo aproximadamente constante, podía servir para ello: las estaciones dieron origen al año, las fases de la Luna a los meses y el ciclo diurno al día. La división de éste en horas tuvo su origen en los relojes solares y la partición de éstas en minutos y, luego, segundos presumiblemente resultó de la confección de relojes de arena y de la cadencia acompasada del corazón humano.

Dado que un segundo es un espacio de tiempo demasiado breve para los seres humanos, su división ulterior abandonó el uso de definiciones relativamente caprichosas y siguió, sin más, el desapasionado imperio del sistema decimal. Esto nos coloca ante una disyuntiva: del mismo modo en que no hay un límite a lo diminuto que puede ser un número, ¿existen intervalos de tiempo arbitrariamente pequeños? Si fraccionamos un segundo dividiéndolo a la mitad, una y otra vez, ¿podremos hacerlo indefinidamente o llegaremos a una unidad mínima e indivisible? Esta pregunta está indisolublemente ligada a otra: ¿existe una distancia mínima entre dos puntos cualesquiera del espacio? La conexión entre ambas cuestiones está dada por la universalidad de la velocidad de la luz en el vacío: si hubiera dos puntos del espacio arbitrariamente cercanos, lo mismo ocurriría con el tiempo ya que la pregunta '¿cuánto tarda la luz en ir de uno al otro?' debería tener una respuesta.

Una pregunta similar se hicieron Demócrito y Leucipo de Mileto en relación a la materia y concluyeron que debía existir una unidad mínima de ésta a la que llamaron átomo. Si estos no existieran podríamos dividir la materia infinitamente y una cucharilla de aceite vertida al mar podría expandirse indefinidamente ya que no habría un límite inferior al espesor de la delgadísima película que, de ese modo, envolvería los océanos. El átomo, en cualquier caso, resultó divisible en constituyentes aún más pequeños, los electrones y el núcleo, y dentro de éste los protones y neutrones. Estos son tan pequeños que la luz demora aproximadamente un yoctosegundo —la cuatrillonésima parte del segundo— en atravesarlos. El mismo tiempo que demora un *quark top* en desvanecerse. Lapsos de tiempo como estos, cuya existencia apenas podemos inferir, son mucho menores que aquellos que se han podido medir directamente, de manera controlada, en un laboratorio y que andan en torno al millón de yoctosegundos.

La jurisprudencia aplicable a preguntas que tengan que ver con las pequeñas escalas es la de la Mecánica Cuántica. Y ésta nos dice que cuanto mayor es la energía que se confiere a un sistema microscópico, más pequeño es el detalle con el que se lo observa; de allí el uso de aceleradores de partículas. En el Gran Colisionador de

Hadrones (LHC) se ha alcanzado una resolución tan fina para la estructura de la materia, que la luz recorrería ese diminuto píxel en una cienmilésima de yoctosegundo. Ninguna máquina fabricada por seres humanos ha inyectado la energía suficiente en un sistema microscópico que permita ir más allá de estas escalas. Pero existen sistemas naturales que, por un mecanismo aún no del todo comprendido, son capaces de acelerar partículas hasta energías millones de veces mayores. Estas partículas recorren enormes distancias en el Universo y eventualmente entran en la atmósfera terrestre: son los llamados rayos cósmicos. El más energético registrado hasta la escritura de estas líneas surcó el espacio experimentando un pixelado que la luz recorrería en unas cienmilmillonésimas de yoctosegundo. Un intervalo de tiempo que nos resulta inimaginable, absurdamente pequeño, y que nos devuelve a la pregunta formulada más arriba, ¿podremos dividir al segundo indefinidamente?

De la legislación del mundo microscópico se desprende el principio de incertidumbre que formuló Werner Heisenberg en 1927. Éste nos dice, entre otras cosas, que mientras mayor resulte la certeza respecto del instante en el que un fenómeno acontece, más grande será la indeterminación de su energía. Y lo más sorprendente es que la Naturaleza saca provecho de ello, permitiendo cierta efervescencia microscópica del vacío que resulta de la continua creación y destrucción de partículas: mientras estos procesos tengan lugar en intervalos de tiempo inferiores a la cienmilmillonésima de yoctosegundo, el principio de conservación de la energía resultará escrupulosamente respetado. Por otra parte, dada la icónica fórmula de Einstein, $E = mc^2$, cuanto más pequeño sea el intervalo temporal observado mayor será la masa de las partículas que se puedan crear espontáneamente en el chispeante vacío.

La posibilidad de determinar un intervalo de tiempo arbitrariamente pequeño, entonces, va inexorablemente de la mano de la disponibilidad ilimitada de energía del vacío. La Teoría de la Relatividad General, por otra parte, nos dice que la acumulación de energía en una región pequeña del espacio da lugar a un agujero negro. Así, si el tiempo pudiera fraccionarse indefinidamente, ¡el Universo estaría infestado de agujeros negros microscópicos! Si la jurisprudencia de la Mecánica Cuántica alcanza a las escalas más diminutas, entonces no puede existir un intervalo de tiempo arbitrariamente pequeño. Una conclusión que choca con la sensación de continuidad en el devenir temporal que experimentamos los seres vivos, enfrentándonos una vez más a la física de la perplejidad que gobierna al universo microscópico. Nada sorprendente si recordamos que nuestros sentidos han sido moldeados por la evolución para desenvolverse en las escalas de tiempo, espacio y materia en las que habitan nuestros cuerpos, alimentos y depredadores.

¿A qué diminuta escala del tiempo es de esperar que la noción de flujo continuo deje de ser una buena aproximación de la realidad? Una pista nos la brindan las constantes fundamentales de la Naturaleza. Estas son cantidades que forman parte de sus leyes y que resultan las mismas en cualquier rincón del Universo observable: la velocidad de la luz, la constante de Newton y la constante de Planck. Cada una de ellas representa la marca de identidad de, respectivamente, la relatividad, la gravedad y la física cuántica. Existe una única combinación aritmética de ellas que da lugar a una escala temporal. No hay otra forma de generar con ellas algo que pueda medirse en segundos. Se la conoce como *el tiempo de Planck* y su propia constitución deja claro que al llegar a esta escala crujirán los cimientos del edificio que sostiene nuestra noción de continuidad temporal.

Si recordamos el valor del instante de tiempo más pequeño que hemos podido medir directamente y de manera controlada en un laboratorio, caben en él tantos tiempos de Planck como horas en la edad del Universo. El tiempo de Planck es extremadamente diminuto, la cientrillonésima parte de un yoctosegundo. Así como en el universo microscópico tenemos dificultades para discernir si los constituyentes de la materia son ondas o partículas, del mismo modo en que su naturaleza corpórea se vuelve elusiva, sabemos que al llegar a la escala de Planck el tiempo, tal como lo entendemos, dejará de existir. Si recortáramos un segundo una y otra vez como si fuera un largo hilo, nos encontraremos que al acercarnos a la escala de Planck la hebra comenzará a desdibujarse, a convertirse en algo completamente irreconocible. ¿Un enjambre de cuerdas microscópicas que vibran? ¿Una enorme colección de bits que de lejos generan la ilusión de un tiempo continuo que fluye? ¿Un píxel de tiempo, como el grano de arena de un reloj? Tal como los átomos y las moléculas son la expresión mínima de la materia, no hay intervalo más fugaz que el tiempo de Planck.

José Edelstein
Doctor en Física
Dpto. Física de Partículas, Universidad de Santiago de Compostela

Capítulo 34
¿Qué es la cuarta dimensión?
(Por Luis Velilla Prieto)

Percibimos el mundo que nos rodea como un espacio de tres dimensiones. Así, para determinar una posición en el espacio necesitamos tres coordenadas espaciales y un punto de referencia para dar esas coordenadas. En nuestra vida cotidiana, utilizamos normalmente sistemas reducidos, como por ejemplo los mapas, que requieren solo dos coordenadas para ubicarnos, ya que damos por hecho que el lugar al que queremos ir está localizado sobre el suelo. Realizamos incluso mayores abstracciones, por ejemplo cuando viajamos, ya que al definir nuestra ruta estamos simplificando el espacio tridimensional a una única coordenada. El localizador GPS [1] nos facilita esa tarea, y solo debemos preocuparnos de una coordenada: la distancia hasta el destino (o desde el punto de partida). Aunque también somos capaces de crear espacios con mayor número de dimensiones, al menos desde el punto de vista teórico, los *hiperespacios*. Estos espacios surgen de manera teórica para intentar resolver problemas físicos, espacios a los que nuestra mente no está acostumbrada, no en vano llevamos millones de años de evolución observando un mundo tridimensional.

Pero dejando atrás dichos hiperespacios, existe algo mucho más cotidiano que conocemos como cuarta dimensión: el *tiempo*. En realidad, no basta con las tres coordenadas espaciales para establecer un evento. No basta con estar en el lugar correcto, también es necesario estar en el momento adecuado en el que un evento ocurrirá. El tiempo es esa dimensión en la que observamos como ocurren las cosas en un determinado orden, como una sucesión de eventos, con un pasado, un presente, y un futuro. Como toda dimensión, requiere de un sistema de medida, un estándar, que sea reproducible y aceptado por todo el mundo. De manera natural, el ser humano comenzó a establecer unidades de medida del tiempo en relación a fenómenos naturales como el día y la noche, las fases de la luna, las estaciones... Fenómenos que ocurren de manera periódica y que se deben al movimiento de nuestro planeta, La Tierra, en el sistema solar. Podríamos usar un péndulo o un reloj de arena para intentar medir el tiempo, pero es evidente que, en general, éstos serían métodos imprecisos que no valdrían como estándares de medida. El segundo, que es la unidad básica de medida del tiempo en los sistemas de unidades más comunes, se definía (hasta 1967) en relación al ritmo de rotación terrestre, y también respecto al período de su órbita alrededor del Sol. En 1967 se establece que estos movimientos son imprecisos como estándar de medida. Desde entonces, el segundo se establece como el lapso temporal que transcurre entre 9.192.631.770 períodos de transición entre dos niveles hiperfinos del isótopo de cesio-133. Es lo que conocemos como *reloj atómico,* cuyo ritmo estamos acostumbrados a escuchar en las emisoras de radio cuando se produce un cambio de hora, como una secuencia de seis pitidos consecutivos, siendo el último (el más largo) el que nos indica el comienzo de la nueva hora. Estos relojes son tan precisos que el mayor desfase que se

produce entre dos relojes atómicos cualesquiera del mundo es de 1 nanosegundo (una milmillonésima parte de un segundo) en un día.

Pero, ¿qué y cómo es en realidad el tiempo? La física trata de darnos una respuesta objetiva para contestar a estas preguntas, y la experiencia nos muestra que el tiempo es relativo y no absoluto. Fue Albert Einstein quien en 1905 ejemplificó esta relatividad del tiempo tratando de describir el movimiento de los objetos. Anteriormente a Einstein, el tiempo se consideraba absoluto, si bien sí se conocía la relatividad del movimiento. Supongamos que tres personas se disponen a realizar un experimento sobre el movimiento, para lo que dos de ellas (A y B) se suben a un tren, mientras que el tercero (C) esperará sentado en la estación observando a las otras dos personas. Supongamos que el tren comienza a moverse, y que A se encuentra sentado dentro del tren y B se encuentra andando a través de uno de los vagones del tren. ¿Cuál es el estado de movimiento de A, B y C? La respuesta dependerá de a quién preguntemos. La persona C nos dirá que A se encuentra en movimiento, mientras que B nos dirá que A está quieto. Por tanto, el estado de movimiento de un objeto es relativo. El tiempo es también relativo, y no me refiero a la percepción subjetiva que cada uno tenemos del tiempo, esa sensación de que las horas se hacen interminables cuando hacemos algo a disgusto. Fuera de esta subjetividad, el tiempo es relativo porque depende del estado de movimiento del observador que trate de medirlo. Mediante las teorías de la relatividad especial y general de Einstein se puede demostrar este hecho, que además se ha probado experimentalmente. Es bastante conocida la llamada *paradoja de los gemelos*, donde se establece que si uno de los hermanos (de una pareja de gemelos) decidiera realizar un viaje desde La Tierra al espacio moviéndose a una velocidad cercana a la de la luz, mientras que el otro gemelo permanece en La Tierra, resultará que a la vuelta del viaje, el gemelo que no viajó habrá envejecido mucho más que el gemelo viajero, fenómeno conocido como *dilatación temporal*. No penséis que este es el secreto de la eterna juventud, pues para cada uno de los observadores el tiempo propio transcurre de manera natural. Existen además pruebas empíricas que demuestran que Einstein estaba en lo cierto, así en 1972 se publicaron en la revista *Science* [2] los resultados sobre el desfase temporal que experimentaban varios relojes atómicos que viajaron en vuelos comerciales en diferentes direcciones (hacia el este y hacia el oeste). No entraremos en las ecuaciones que explican este fenómeno, pero este experimento reproducía satisfactoriamente las predicciones de Einstein, de modo que cuando los relojes viajaban hacia el este el lapso de tiempo que medían era menor que cuando viajaban hacia el oeste, ya que en realidad la velocidad (que mediría un observador fuera de La Tierra, como el observador C de nuestro ejemplo del tren) en ambos viajes no será exactamente igual debido a la rotación terrestre (que rota hacia el este). Lo cierto es que para las velocidades de los aviones comerciales y la duración de estos viajes, los desfases que se producen entre los relojes son del orden de las decenas a pocos cientos de nanosegundos, por lo que en nuestra vida cotidiana no percibiríamos este desfase aunque pasáramos el resto de nuestra vida viajando en avión. Sin embargo, si que tenemos en cuenta este efecto cuando realizamos cálculos precisos, como por ejemplo con la mencionada tecnología GPS.

De la teoría de la relatividad especial de Einstein se desprende que espacio y tiempo van ligados, como hemos dicho, por el estado de movimiento. Desde entonces se trata el espacio-tiempo como una entidad indivisible, de cuatro dimensiones, tres espaciales y una temporal. El espacio-tiempo es el tejido que conforma el Universo, y que además de verse afectado por el estado de movimiento de quién lo observa, cambia por la presencia de objetos masivos, como indica la teoría de la relatividad general de

Einstein. Para entender cómo funciona este efecto, propongamos otro experimento. Sujetemos una tela (por ejemplo una toalla) por sus cuatro esquinas con la ayuda de otra persona, aplicando cierta tensión. La toalla será nuestro espacio-tiempo, y cualquier cosa que ocurra en este Universo, ocurrirá sobre la superficie de la toalla. Si aplicando la misma tensión pedimos a una tercera persona que coloque un objeto pesado (por ejemplo una pelota) en el centro, ocurrirá que la toalla se curvará hacia el punto donde la pelota se apoya. La forma de nuestro espacio-tiempo ha cambiado por la presencia de un objeto masivo. Así que el contenido del Universo influye en el espacio-tiempo de éste, y además afecta a sus propiedades. De hecho en nuestro Universo, los objetos muy masivos como los agujeros negros son capaces de desviar e incluso “tragar” las partículas que componen la luz (los fotones) y que viajan a una velocidad aproximada de 300,000 kilómetros por segundo. Los agujeros negros pueden curvar tanto el espacio-tiempo que ni los fotones que pasan cerca de ellos son capaces de evitar caer en ese pozo. Probad a hacer rodar una canica por la superficie de vuestra toalla e intentad que la canica no caiga en el pozo creado por la pelota. Si la pelota llega a curvar mucho la toalla llegará un punto en que no tendréis fuerza suficiente para evitar que la canica caiga en el pozo, salvo que escojáis una trayectoria adecuada. Esta teoría de la relatividad general predice además que el tiempo transcurre de manera más lenta para un observador que se encuentre cerca de un objeto muy masivo (bajo la influencia de su campo gravitatorio). Este efecto también se ha probado experimentalmente, gracias a los experimentos mencionados anteriormente sobre los relojes atómicos viajeros [2,3]. Si situamos dos relojes idénticos uno sobre la superficie de La Tierra y el otro reloj a una altura determinada sobre su superficie, resultará que el reloj situado en la superficie se retrasa comparado con el otro. Si bien el desfase producido sería de unos pocos segundos si los dejáramos funcionando millones de años. La razón del desfase es que el reloj que se encuentra sobre la superficie terrestre experimenta una atracción gravitatoria debida a La Tierra mayor que el reloj en altura.

Las matemáticas que explican estos fenómenos son realmente complejas, pero tanto teoría como experimentación muestran que el tiempo es relativo y que depende del estado de movimiento del observador que lo mide, así como de la materia que constituye el Universo, ya que los campos gravitatorios creados por esa materia alteran el propio espacio-tiempo. No obstante, lo que sabemos sobre el espacio-tiempo lo acabamos de aprender, por así decirlo, pues las teorías de la relatividad tienen solo cien años. Al igual que hemos sustituido los relojes de sol por los atómicos, en el futuro nuestra concepción del tiempo será distinta. Si queréis llegar a conocer el futuro, la física os recomienda viajar lo máximo posible.

Notas:
[1] Del inglés Global Positioning System, o sistema de posicionamiento global.
[2] “Around-the-world atomic clocks: observed relativistic time gains”. J.C. Hafele, Richard E. Keating. Science, 14 de Julio de 1972.Vol. 177, Issue 4044, pp. 168-170.
[3] “Optical clocks and relativity”. C.W Chou, D.B. Hume, T. Rosenband, D.J. Wineland. Science, 24 de Septiembre de 2010. Vol. 329, Issue 5999, pp. 1630-1633.

Luis Velilla Prieto
Doctorando Astrofísica
Instituto de Ciencia de Materiales de Madrid (CSIC)

Capítulo 35
¿Qué quiere decir que el espacio está curvado?
(Por César González Arranz)

«Si he logrado ver más lejos, ha sido porque he subido a hombros de Gigantes». Newton escribió estas palabras a su colega Robert Hooke en una carta fechada el 15 de febrero de 1676. En ella, ambos genios británicos firmaban la paz a petición de la Royal Society, de la que los dos eran miembros, después de que hubiesen mantenido una disputa científica. Isaac Newton (1642 – 1727) ha sido una de las figuras clave en el intento del hombre por hallar una explicación científica del Universo. Sus campos de investigación abarcan temas de óptica, matemáticas, astronomía, mecánica celeste, etc. Pero por lo que más se conoce a Newton en ámbitos no profesionales es por su famosa teoría de la Gravedad; cualquier persona sabe que los objetos caen a la superficie terrestre debido a la fuerza gravitatoria que ejerce nuestro planeta sobre todos ellos, y que fue Newton quien lo explicó. La interpretación más tradicional de la frase de Newton es que la Ciencia se construye poco a poco gracias a que el trabajo de unos científicos es utilizado posteriormente por otros para seguir avanzando y progresando. Por cierto… la frase no es de Newton sino que se atribuye al filósofo neoplatónico Bernardo de Chartres, del siglo XII.

La cita de Bernardo de Chartres puede aplicarse al caso que nos ocupa en este capítulo y, cómo no, Newton está implicado: él vio más al subirse a hombros de gigantes, pero científicos posteriores a su persona, también. Para Newton, el espacio y el tiempo eran absolutos. La realidad física consistía en un espacio compuesto por tres dimensiones espaciales (alto, ancho y largo) y un tiempo que solo podía recorrerse en un único sentido, del pasado hacia el futuro. Que fuesen absolutos significaba que siempre, en todo momento y lugar, existiría un patrón igual de espacio y de tiempo común a todos los observadores en cualquier situación física. El tiempo marcado por un determinado reloj sería el mismo para todos los lugares del espacio.

Por otro lado, los cuerpos físicos (un planeta, una piedra, cualquier objeto en definitiva) presentes en el espacio se representaban por masas puntuales con unas coordenadas espaciales muy determinadas en cada instante de tiempo (alto, ancho y largo). Estas masas se atraerían unas a otras mediante una fuerza de atracción que Newton denominó fuerza de la Gravedad, y que era de acción instantánea, es decir, que el cambio de posición de dos masas debida a sus fuerzas de gravedad sucedía instantáneamente, sin que transcurriese intervalo de tiempo absoluto entre ambos sucesos. Este hecho perturbaba a Newton… ¿cómo podía una fuerza ser de acción inmediata y, además, actuar a distancia? La solución tardaría en llegar casi dos siglos. El movimiento de estos cuerpos venía regido por las leyes de la dinámica que Newton había formulado, pero solo si se encontraban en un sistema de referencia que permaneciesen en reposo o moviéndose a velocidad constante con respecto al sistema de referencia absoluto. A estos sistemas se les denominaba inerciales y en ellos las

ecuaciones del movimiento eran las mismas. Si no era el caso, es decir, si se movían variando la velocidad, con una aceleración, con respecto del sistema de referencia absoluto, entonces aparecían en las ecuaciones dinámicas de Newton unas cantidades que dependían de la aceleración, llamadas fuerzas de inercia.

Recapitulando hasta ahora, tenemos que para Newton existía un espacio absoluto, con un tiempo absoluto igual para todos los objetos existentes en el espacio. Estos se movían unos con respecto a otros, y con respecto al espacio absoluto, debido a sus interacciones gravitatorias siguiendo las leyes de la dinámica de Newton. Si estos cuerpos se movían con velocidad constante con respecto al espacio absoluto, las ecuaciones newtonianas del movimiento no variaban si pasábamos entre ambos sistemas de referencia: eran sistemas de referencia inerciales.

Pero si realmente deseamos saber qué quiere decir que el espacio está curvado, necesitamos avanzar dos siglos más allá de la época de Newton y fijarnos en el trabajo del escocés James Clark Maxwell (1831 – 1879) sobre el electromagnetismo. Maxwell también subió a hombros de gigantes ya que dedujo cuatro ecuaciones con las que describir perfectamente todos los fenómenos de la electricidad y el magnetismo usando los resultados experimentales obtenidos a lo largo de muchos años por científicos anteriores a él como Coulomb, Gauss, Ampere, Faraday y otros muchos más. ¿Qué tienen que ver las ecuaciones del electromagnetismo de Maxwell con las ecuaciones de la dinámica de Newton? Pues que si las ecuaciones del movimiento newtonianas no variaban si pasábamos entre dos sistemas de referencia inerciales, se originaba la duda de si ocurriría lo mismo con las ecuaciones del electromagnetismo, es decir, que si teníamos dos sistemas de referencia inerciales y pasábamos de uno a otro, las ecuaciones de Maxwell tampoco variarían. En un principio, teóricamente, al aplicar la transformación necesaria para pasar de un sistema al otro (inercial), las ecuaciones de Maxwell no quedaban igual… pero después de numerosos experimentos se comprobó que las ecuaciones del electromagnetismo no sufrían variación y que, por consiguiente, los fenómenos electromagnéticos se exhibían exactamente igual en cualquier sistema de referencia inercial. Este principio de relatividad se cumplía tanto con las ecuaciones de la mecánica clásica como con las del electromagnetismo, aun siendo tan diferentes unas de las otras… la solución vino de la mano del físico alemán Albert Einstein (1879 – 1855). El principio de relatividad lo dividió en dos: la Teoría Especial de la Relatividad (1905) y la Teoría General de la Relatividad (1915).

La Teoría Especial de la Relatividad (TER) se ocupa de los movimientos rectilíneos con velocidad constante. Todos los resultados que se obtienen de la TER se desprenden de un postulado básico: la velocidad de la luz es constante, independientemente del sistema de referencia en el que se mida y, además, es inalcanzable: ningún cuerpo material puede sobrepasarla, ni siquiera acercarse a ella. De una manera muy resumida, diremos que la TER define al tiempo y al espacio como algo no absoluto. Personas que estén viajando en, digamos, naves espaciales que están moviéndose a velocidad constante unas con respecto a las otras (son sistemas inerciales) medirán un tiempo determinado, diferente para cada una de ellas, que se denomina tiempo propio. No existe, por tanto, un tiempo absoluto igual para todas y cada una de las naves. De igual manera, el espacio tampoco es absoluto: un tripulante de una nave, número uno, medirá la longitud de otra nave, número dos, y obtendrá un resultado diferente al que saque el tripulante de una tercera nave que también mida el tamaño de la segunda. ¡Pero es que el tamaño que mida el tripulante de la nave dos de su propia nave tampoco coincidirá con el de la nave uno ni con el de la nave tres! ¿Cuál

es la longitud real única y verdadera de la nave dos? ¡Ninguna! Porque todas son verdaderas, la longitud dependerá del sistema de referencia desde el que se la mida. Cuanto más cercano a la velocidad de la luz (c = 300 000 km/s) se mueva un sistema viajando a velocidad constante con respecto a uno en reposo, más se notarán las diferencias en la medida del tiempo y longitud. Desde el sistema en reposo se verá cómo el tiempo en el sistema en movimiento transcurrirá más lentamente que el tiempo en el sistema en reposo y comprobará cómo las longitudes en el sistema que está moviéndose serán más cortas que las del sistema en reposo: dilatación del tiempo y contracción de longitudes.

Ya nos vamos acercando, por fin, a la pregunta propuesta en este capítulo: ¿qué quiere decir que el espacio está curvado? La explicación reside en la Teoría General de la Relatividad (TGR), que se ocupa de los movimientos acelerados (recordemos que la TER se preocupaba de los movimientos rectilíneos a velocidad constante). Antes de seguir, hacer notar que el movimiento acelerado a gran escala más significativo es el causado por los objetos sujetos a un campo gravitatorio. Antes que Einstein, el matemático ruso Hermman Minkowski (1864 – 1909) se percató de que jugando con las ecuaciones de la TER podía conseguir tratar las tres coordenadas espaciales (alto, ancho y largo) y el tiempo no como algo independiente sino como una especie de entidad compuesta por cuatro cantidades íntimamente relacionadas entre sí, las tres coordenadas espaciales y la coordenada de tiempo, y que formarían una geometría no euclidiana (la suma de los ángulos de un triángulo no sumarían 180º, como en una geometría sí euclidiana) en un espacio de cuatro dimensiones: el continuo espacio-tiempo. Einstein usó las conclusiones obtenidas por Minkowski para crear una nueva teoría acerca de la fuerza gravitatoria. Básicamente, esta teoría dice que cualquier masa presente en el Universo (un planeta, un satélite, una estrella, una galaxia, etc.) causa una deformación en el continuo espacio-tiempo, y es esta deformación lo que denominamos «gravedad». De manera más simple, se puede definir el continuo espacio-tiempo como el lugar tetradimensional en el que están colocados todos los astros en el Universo, y lo podemos representar, figuradamente, como si fuera una especie de malla elástica (esta es una manera de simplificar muchísimo las cosas, pero no hay otra manera porque el espacio-tiempo tiene 4 dimensiones, como ya hemos indicado, imposible de visualizar por nosotros pero perfectamente definible matemáticamente). Estas cuatro coordenadas son las 3 espaciales (largo, ancho y alto) y la del tiempo. Imaginad que en la malla colocamos ahora un balón de fútbol, que va a representar a un planeta: ¿qué hace el balón en la malla? La deforma. Pues de la misma manera, el planeta deforma el espacio–tiempo, causando una «curvatura»… aquí tenemos la explicación de que el espacio esté «curvado». Cuanto mayor sea la masa del planeta, mayor será la deformación (la curvatura), y más intensa la «fuerza de gravedad». Si lanzamos una canica por la malla en dirección al balón, esta canica va a acabar chocando con el balón porque va a caer por la pendiente de la malla que se crea al deformarse esta última por la presencia del balón. De igual manera, si enviamos una nave tripulada hacia el planeta, cuando esta llegue a la zona del espacio deformada por el planeta ¿qué va a sentir el astronauta que va en su interior? Pues que hay una especie de fuerza (de la «gravedad») que proviene del planeta y que lo atrae hacia él; si no pone remedio, su nave chocará con dicho planeta. Pero fijaos que no existe ninguna fuerza (gravitatoria), lo único que ocurre es que la nave sigue la forma del espacio-tiempo, que está deformado por la presencia del astro. Lo mismo que le sucede a la nave de nuestro ejemplo le pasa a un rayo de luz; la luz viaja por el espacio siguiendo la geometría del espacio–tiempo, si este es curvo, la trayectoria de la luz es curva. Esta era una de las predicciones de la TGR que se comprobó experimentalmente en el año 1919 cuando se observó durante un

eclipse total de Sol que la posición de las estrellas cercanas al Sol había cambiado ligeramente debido a la presencia de nuestra estrella, tal y como había predicho Einstein.

Desde la época de Newton, hemos pasado de tener un espacio y un tiempo absolutos a tener un tiempo y un espacio propios del sistema de referencia al que pertenecen. A tener una fuerza gravitatoria a distancia y de efecto inmediato a un tejido espaciotemporal de cuatro dimensiones cuya curvatura es la que provoca lo que llamamos «gravedad». Las ecuaciones que conforman la TGR pueden aplicarse a muchos supuestos físicos, incluso al Universo como un todo… lo más fascinante es que de esta manera puede llegarse a conocer el origen, evolución y destino del Universo. Así nació la Cosmología. Por otro lado, muchas han sido las predicciones que la TGR dejó con su formulación. Todas y cada una de ellas se han ido cumpliendo poco a poco. La más reciente, la detección de ondas gravitacionales en 2015 a cargo del experimento Advanced LIGO (Advanced Laser Interferometer Gravitational-Wave Observatory). Imaginad un objeto muy compacto y masivo, como un agujero negro o una estrella de neutrones. Si estos objetos se mueven por la malla del espacio tiempo a altas velocidades, la TGR dice que generarán en este último unas ondulaciones parecidas a las que se producen en un estanque cuando tiramos una piedra al agua. Las ondas gravitacionales son, pues, ondulaciones en el propio tejido del espacio-tiempo que se generan en procesos en los que hay involucrada una enorme cantidad de energía. Se abre una nueva manera de estudiar el Universo que nos rodea, que hasta este momento permanecía oculta… la astronomía de ondas gravitacionales… pero esto ya es otro capítulo…

Bibliografía:
“Espacio y tiempo”, Teodoro Vives. Equipo Sirius (2006)
“Sobre la teoría de la relatividad especial y general”, Albert Einstein. Alianza editorial (2003)
“La geometría del universo”, Manuel de León. CSIC, los libros de la catarata (2012). Colección ¿qué sabemos de?

César González Arranz
Técnico Superior en Sistemas de Telecomunicación e Informáticos.
Planetario de Madrid

Capítulo 36
¿En qué consiste la Teoría de la Relatividad Especial o Restringida de Einstein? ¿Por qué nada puede moverse más rápidamente que la velocidad de la luz?

(Por Julio Gutiérrez Muñoz)

36.1 Introducción

La idea de relatividad del movimiento no es nueva, de hecho, Galileo le dedicó unos maravillosos párrafos en sus *Diálogos*. No obstante, sistemáticamente incluso hoy, los libros de texto han obviado el enunciado del Principio de Relatividad, esencial en la Mecánica Clásica, lo cual ha proporcionado no pocos quebraderos de cabeza a los estudiantes. El Principio de Relatividad mantiene su vigencia en la Física actual y, como veremos, la única diferencia entre los enunciados de Galileo y de Einstein reside en la velocidad de transmisión de la información de las interacciones entre los cuerpos –mientras para Galileo es infinita, para Einstein es finita y tiene como valor máximo la velocidad de la luz en el vacío–. Las consecuencias de esta, aparentemente, simple diferencia han provocado cambios de enorme importancia para la Física a partir de finales del siglo XIX.

Advertimos al lector que en este capítulo va a encontrarse con expresiones y demostraciones matemáticas, no obstante, son extremadamente sencillas y propias de un ejercicio algebraico elemental. Sin ellas, la extensión del texto sería demasiado larga.

36.2 Sistemas de Referencia y Principio de Relatividad

Antes de entrar en materia, debemos recordar en qué consiste un Sistema de Referencia, pues sin una idea clara de su concepto difícilmente entenderemos cómo se plantea la relatividad de Einstein, tanto la especial correspondiente al presente capítulo, como la general que analizaremos en el siguiente; es más, sería prácticamente imposible explicar en qué consiste una aceleración ficticia como la debida al efecto Coriolis que estudiaremos en el capítulo 70. Además, es muy frecuente, incluso en buenos libros de texto, confundir sistema de referencia y sistema de coordenadas, y creemos necesario hacer un pequeño paréntesis para clarificar sus diferencias: mientras los primeros dan respuesta a la necesidad dinámica de descripción de los fenómenos de interacción (los únicos que pueden producir cambios en la Naturaleza), los segundos simplemente utilizan la geometría y el cálculo matemático para describir cuantitativamente las consecuencias de dichos fenómenos.

Definimos Sistema de Referencia –en adelante abreviado por* SR*– como el laboratorio en el que se hacen las observaciones de los cambios provocados por las interacciones fundamentales de la Naturaleza, siempre en relación con el movimiento. Nuestro planeta es un sistema de referencia en el que estudiamos los fenómenos más variados, incluidos los complicados movimientos galácticos. No

obstante, en el mismo momento en que intentemos medir la velocidad de un objeto, notaremos que su estado de movimiento depende de la propia velocidad del observador que hace la medida. Podríamos estudiar un objeto en caída libre desde la superficie terrestre y le veríamos dotado de velocidad y aceleración, pero también podríamos abandonarnos como observadores a una caída libre paralela para estudiar el mismo fenómeno y, en ese caso, el objeto en cuestión tendría velocidad y aceleración nulas respecto de nosotros. Intuimos, por tanto, que las magnitudes cinemáticas reposo, velocidad y aceleración son totalmente relativas al observador.

Podemos añadir un sistema de coordenadas –en adelante abreviado por SC–, a fin de poder cuantificar las posiciones relativas y las magnitudes cinemáticas. En el ejemplo de la Tierra como SR, nos puede servir de SC el formado por las distancias a su centro y los dos ángulos determinantes de la latitud y longitud sobre la superficie terrestre. Obviamente, existen infinitos sistemas de coordenadas para un SR dado, pues infinitos son los puntos de origen posibles. Es más, se pueden tomar diferentes sistemas matemáticos de medida de posiciones, velocidades y aceleraciones: coordenadas polares, esféricas, cilíndricas, cartesianas, etc. En cualquier caso el SR es siempre el mismo: la Tierra.

Ahora bien, ¿las aceleraciones no son debidas a las fuerzas? Entonces, ¿cómo es posible que sean relativas al observador y no sean magnitudes absolutas? Es aquí donde el Principio de Relatividad nos permite eludir esta dificultad conceptual.

En un SR con velocidad constante es muy fácil describir los fenómenos derivados de las interacciones fundamentales. Sin embargo, si el SR está acelerado, las cosas se complican hasta el punto de tener que introducir fuerzas ficticias –las **fuerzas de inercia**– para explicar los aparentes fenómenos inexistentes, debidos exclusivamente a la aceleración del sistema (véase el epígrafe 37.2 en el próximo capítulo). Por estas razones, distinguimos dos tipos de sistemas de referencia:

a) Sistemas de Referencia Inerciales, **SRI**, son aquellos que se mueven con velocidad constante tanto en módulo como en dirección y, por supuesto, sentido.

b) Sistemas de Referencia No Inerciales, **SRNI**, son aquellos que están dotados de cualquier tipo de aceleración.

En la Física Clásica, un SRI es el esquema más idóneo para estudiar la dinámica de cualquier sistema físico pues, sea cual sea el estado de movimiento del observador, las fuerzas y las aceleraciones medidas son siempre las mismas. Podemos enunciar el Principio de Relatividad en la forma: ***"Todos los sistemas de referencia en movimiento relativo de traslación rectilínea y constante –sistemas de referencia inerciales– son equivalentes desde el punto de vista mecánico; es decir, los experimentos mecánicos se desarrollan de igual manera y las leyes de la Mecánica son las mismas para todos ellos"***.

Surge una primera pregunta obvia: ¿qué criterio se utiliza para elegir un determinado SR? La elección es más bien una cuestión de comodidad, cuantos más objetos o partículas tengamos en reposo mejor. En la superficie de la Tierra muchos objetos están en reposo o se mueven con velocidades constantes, por consiguiente, no será una buena elección de SR la Estación Espacial Orbital, con movimiento relativo acelerado respecto de la superficie terrestre.

La segunda pregunta es inmediata: ¿existe algún SR privilegiado que posibilite definir magnitudes cinemáticas de forma absoluta? La relatividad de Galileo permite definir tal sistema, y eso es precisamente lo que se buscaba a finales del siglo XIX, cuando se pensaba que el Éter era la esencia espacial del Universo y todo, incluida la luz, tendría un movimiento relativo a dicha entidad. Los experimentos de Michelson y Morley demostraron que no existía tal posibilidad, pues la luz no tenía movimiento relativo alguno respecto del Éter, y la Física, y con ella nuestra concepción de las magnitudes dinámicas y de la naturaleza del Universo, dieron un vuelco. Ello llevó a postular a Einstein el Principio de Relatividad de una manera más general y universal a como lo había hecho Galileo: ***"Todos los sistemas de Referencia Inerciales son equivalentes. No se puede realizar ningún experimento para determinar la velocidad absoluta de un SRI porque solo podemos medir velocidades relativas de unos sistemas respecto de otros. Las leyes de la Física han de ser idénticas, sin distinción alguna, en cualquier SRI"***.

Este enunciado tiene un añadido para dar cuenta de los resultados de Michelson y Morley: ***"Sea cual sea el observador que mida la velocidad de la luz en el vacío encontrará siempre el mismo valor"***. Esta afirmación es una conclusión lógica del abandono de la idea de la existencia del Éter e implica que el espacio es inherente al sistema de referencia, cada SR tiene su "propio" espacio por el que se propaga la luz. Veamos el porqué. Cualquier onda sea cual sea su naturaleza, tiene una velocidad de propagación debida a la naturaleza del medio por el que se propaga y no puede depender de la velocidad relativa entre el emisor y dicho medio. Ello debe ser así para preservar la causalidad. Efectivamente, imaginemos la colisión entre dos vehículos, uno parado y el otro animado de gran velocidad. Si el sonido se propagara a velocidad dependiente de la velocidad relativa entre el emisor y el aire, un observador alejado oiría primero el ruido producido por el vehículo veloz y después percibiría el ruido del destrozo del vehículo en reposo, o a la inversa, según que el vehículo en movimiento se acercara o alejara del observador. Trasládese esta paradoja a cualquier interacción, en particular la electromagnética. Otra cosa es que el medio se mueva respecto del observador; en caso de viento, el sonido se propaga a sus 340 m/s más la velocidad del viento relativa al observador. Por consiguiente, dado que la luz no necesita medio alguno para propagarse, no puede estar afectada por el movimiento de algo que no existe.

Algún lector podrá pensar que sí hay una forma de distinguir quien se mueve respecto de quien, y es utilizando el efecto Doppler que da valores diferentes de las frecuencias de las ondas dependiendo de cuál de los dos se mueva respecto del medio, el emisor o el receptor. Esto es verdad para las ondas que necesitan de un medio para propagarse, pero la luz no puede dar resultados diferentes al no necesitar medio de propagación y, precisamente por ello, el efecto Doppler electromagnético es simétrico.

36.3 Consecuencias del Principio de Relatividad: relatividad especial o restringida

En primer lugar, vamos a definir un par de conceptos utilizados profusamente en la literatura dedicada a la relatividad einsteniana. Llamaremos **suceso** a cualquier fenómeno que tiene lugar en un instante de tiempo y un lugar determinados. Dos sucesos serán **simultáneos** cuando ocurran en lugares distintos pero en el mismo instante de tiempo.

Tomemos dos SRI que se mueven uno respecto del otro a velocidad constante V, uno consiste en una plataforma de longitud *L* –un tren, nave espacial, meteorito, etc.– que llamaremos SRI2, y el otro, SRI1, es un observador que consideramos en reposo para facilitar nuestro razonamiento. Podemos definir un sistema de coordenadas cartesianas de forma que V tenga la dirección del eje X con sentido positivo, en ambos sistemas. Las figuras 1 y 2 ilustran la situación.

Justo en el centro de la plataforma existe un emisor de luz que envía simultáneamente destellos a derecha e izquierda. Suponemos que sincronizamos los relojes en ambos sistemas de referencia justamente cuando se dispara el emisor central de la plataforma. En los extremos se sitúan dos receptores que levantan un banderín a la llegada de los fotones. Analicemos los sucesos en los dos SRI, suponiendo que la propagación se hace en el vacío, pero sin olvidar que ambos observadores medirán la misma velocidad para la luz, aproximadamente, c = 300 000 km/s.

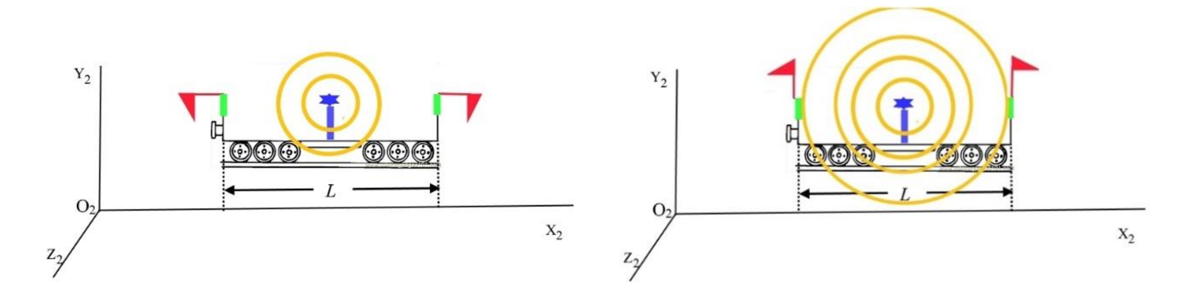


Figura 1: *En el SRI2, el observador está en reposo respecto del emisor y los receptores de luz. Como la luz no puede tener velocidad relativa respecto del espacio, en este sistema las esferas que representan los frentes de onda luminosos están centradas en el emisor.*

SRI2, observador O2

Suceso 2₁: Envío en el instante t_1 de sendos destellos luminosos a derecha e izquierda.

Suceso 2₂: Recepción simultánea de los rayos luminosos por los sensores de los extremos en el instante t_2.

Dado que en SRI2 todo está en reposo, la distancia recorrida por los dos destellos es la misma, e igual a L/2, y el valor del intervalo de tiempo invertido será:

$$\Delta t \;=\; t_2 - t_1 = \frac{L/2}{\mathrm{c}} \qquad (36\text{-}1)$$

SRI1 Observador O1

Suceso 1₁: Observación de la salida simultánea de ambos rayos de luz del centro de la plataforma, en el instante t_1.

Debemos notar que, en este sistema de referencia, ambos rayos no recorren la misma distancia, debido al movimiento relativo, a velocidad V, entre los observadores. El rayo que parte hacia la izquierda debe recorrer una distancia:

$$D_1 \;=\; L/2 \;-\; \mathrm{V}\,\Delta t_1 \qquad (36\text{-}2)$$

siendo Δt_1 el tiempo invertido por el rayo de luz en llegar al extremo izquierdo del vagón y V Δt_1 la distancia recorrida hacia la derecha por dicho extremo que, durante Δt_1, se acerca al punto de emisión.

D_1 es menor que la distancia D_2 recorrida por el rayo que parte hacia la derecha:

$$D_2 = L/2 + \mathrm{V}\,\Delta t_2 \qquad (36\text{-}3)$$

siendo Δt_2 el tiempo invertido por el rayo de luz en llegar al extremo derecho del vagón y V Δt_2 la distancia recorrida hacia la derecha por dicho extremo que, durante Δt_2, se aleja del punto de emisión.

El lector habrá notado que hemos supuesto intervalos de tiempo diferentes, lo cual es cierto, como veremos enseguida, por causa de que O1 también mide una velocidad de la luz exactamente igual a: c = 300 000 km/s.

Suceso $\boldsymbol{1_2}$: Llegada de los fotones al sensor izquierdo y levantamiento del banderín. El tiempo invertido se expresa:

$$\Delta t_1 = \frac{D_1}{c} = \frac{L/2 - \mathrm{V}\,\Delta t_1}{c} \qquad (36\text{-}4)$$

Un sencillo ejercicio de álgebra nos lleva al valor:

$$\Delta t_1 = \frac{L/2}{\mathrm{V} + \mathrm{c}} \qquad (36\text{-}5)$$

Suceso $\boldsymbol{1_3}$: Llegada de los fotones al sensor derecho y levantamiento del banderín. El tiempo invertido se expresa:

$$\Delta t_2 = \frac{D_2}{c} = \frac{L/2 + \mathrm{V}\,\Delta t_2}{c} \qquad (36\text{-}6)$$

El mismo sencillo ejercicio de álgebra anterior proporciona:

$$\Delta t_2 = \frac{L/2}{\mathrm{c} - \mathrm{V}} \qquad (36\text{-}7)$$

Resulta evidente que las expresiones (36-1), (36-5) y (36-7) no dan el mismo valor para los tiempos que invierte la luz en llegar a los sensores, los cuales dependen de qué observador realice la medida. Mientras que en el SRI2 los tiempos de llegada a los sensores derecho e izquierdo es el mismo, y ambos banderines se levantan simultáneamente, en el SRI1, la luz llega antes al sensor izquierdo, en un tiempo inferior al medido por O2, y tras un tiempo mayor, incluso al medido por O2, se levantará el banderín derecho. ¿Qué ha pasado?

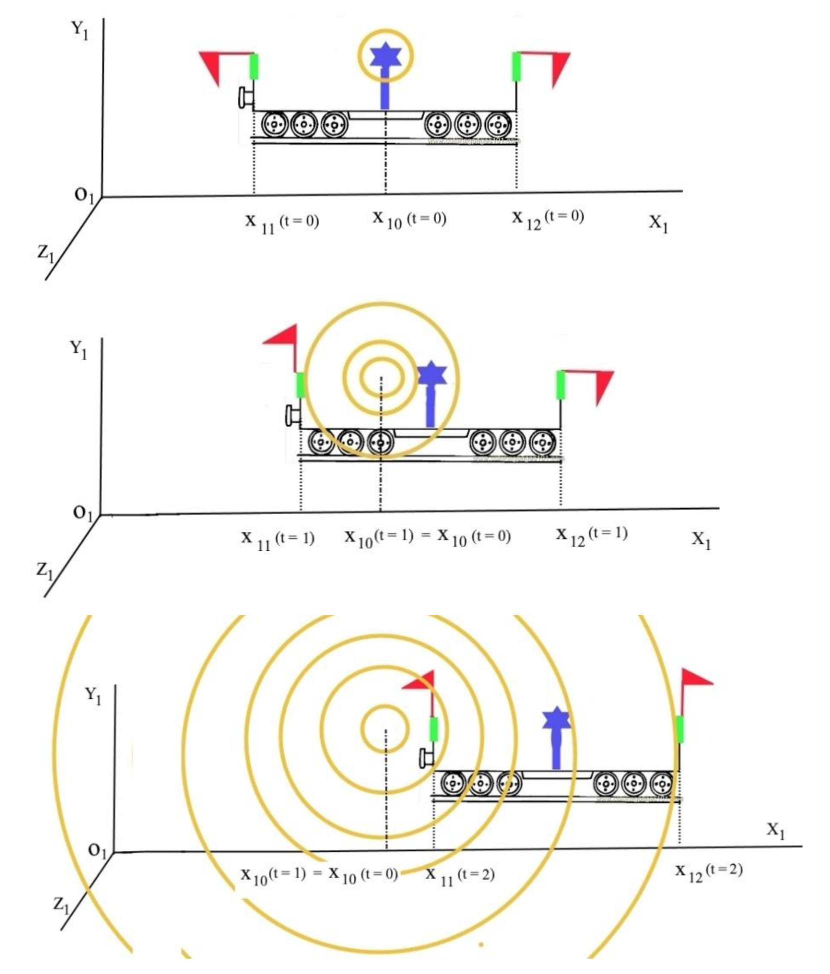


Figura 2: *En el SRI1, el observador está en movimiento relativo respecto del emisor y los receptores de luz. Como la luz no puede tener velocidad relativa respecto del espacio, y el espacio es inherente al SR, en este sistema las esferas que representan los frentes de onda luminosos están centradas en el punto en que se emitió la luz, es decir, la posición del emisor en el instante inicial.*

Como consecuencia de la invariancia de la velocidad de la luz para todo observador inercial, los intervalos de tiempo son relativos al observador y, más aún, sucesos que son simultáneos en un sistema de referencia no lo serán en otro que se mueva respecto del anterior, es decir, el tiempo es también inherente al sistema de referencia.

En la relatividad de Galileo los intervalos de tiempo son invariantes debido a que la velocidad de la luz se considera infinita. De hecho, a velocidades V muy pequeñas frente a c, los intervalos de tiempo (36-1), (36-5) y (36-7) son prácticamente idénticos. Sin embargo, cuando V se aproxima a c, las diferencias pueden ser muy grandes. Supongamos una "plataforma" de longitud L = 540 000 km, que se mueve a velocidad de 240 000 km/s respecto a O1, los resultados de los valores de los intervalos de tiempo serían: a) en O2 idénticos e iguales a 0,9 segundos; b) en O1 diferentes y con un valor de 0,5 segundos para el sensor izquierdo y 4,5 segundos para el tiempo entre la emisión del rayo de luz y el levantamiento del banderín derecho.

El fenómeno no es un resultado absurdo o debido al experimento propuesto, de hecho, cualquier reloj que se vea en movimiento marcará un tiempo más lento que si se observa en reposo. Las figuras 3 y 4 esquematizan lo que ocurre para un reloj que haga "tic" cada vez que un rayo de luz sale de un emisor y haga "tac" cuando los fotones llegan a un receptor, tras haberse reflejado en un espejo, siguiendo un camino vertical de altura H. En un SRI2, en el cual el reloj está en reposo, el tic-tac tiene un valor (véase la figura 3):

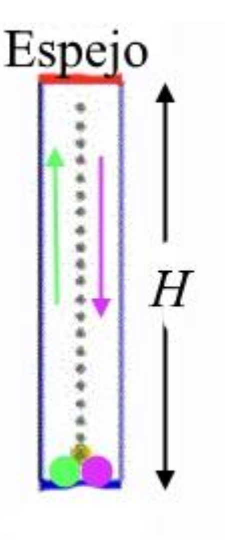


Figura 3: *En un SRI en el que se encuentra en reposo un reloj luminoso que hace tic tac a la emisión y recepción de un rayo de luz, la distancia recorrida por la luz es igual a 2H.*

$$\Delta t_{reposo} = \frac{2H}{c} \qquad (36\text{-}8)$$

En un SRI1, en el cual el reloj se mueve a velocidad V, la luz debe recorrer entre el tic y el tac una distancia (véase la figura 4):

$$D = 2\sqrt{\frac{V^2\ \Delta t^2_{movimiento}}{4} + H^2} \qquad (36\text{-}9)$$

Ello supone un intervalo de tiempo de tic-tac:

$$\Delta t_{movimiento} = \frac{D}{c} = \frac{2\sqrt{\frac{V^2\ \Delta t^2_{movimiento}}{4} + H^2}}{c} \qquad (36\text{-}10)$$

De nuevo, un sencillo ejercicio de álgebra nos lleva al valor:

$$\Delta t_{movimiento} = \frac{2H}{c}\frac{1}{\sqrt{1-\frac{V^2}{c^2}}} = \frac{\Delta t_{reposo}}{\sqrt{1-\frac{V^2}{c^2}}} \qquad (36\text{-}11)$$

Evidentemente (36-11) da un resultado mayor que (36-8), tanto mayor cuanto más se aproxime V a c, tendiendo a infinito cuando la velocidad relativa sea la propia velocidad de la luz. Esta es una de las razones por las cuales la velocidad de la luz no se puede sobrepasar.

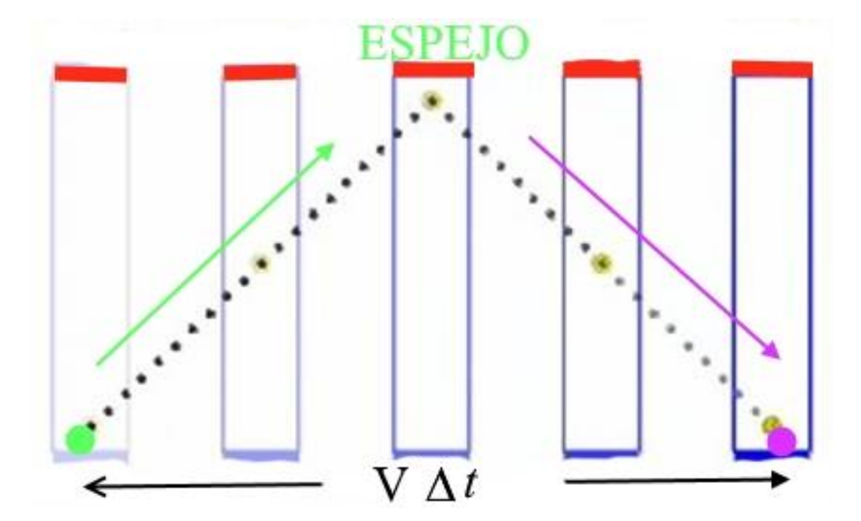


Figura 4: *En un SRI, en el que se mueve a velocidad* V *un reloj luminoso que hace tic tac a la emisión y recepción de un rayo de luz, la distancia recorrida por la luz es mayor que 2H y depende del valor de* V.

Las ecuaciones que hacen el cambio de coordenadas para las posiciones, velocidades y aceleraciones entre dos sistemas de referencia que se mueven a velocidad relativa V, reciben el nombre de **Transformaciones de Lorenz**. De ellas se puede extraer otra consecuencia, que no vamos a demostrar para no agrandar demasiado este capítulo, consistente en que las distancias entre dos puntos en la dirección del movimiento se acortan –lo contrario de lo que le sucede al tiempo–, precisamente con el mismo factor raíz cuadrada que aparece en (36-11), expresión repetida sistemáticamente en cualquier ecuación de la relatividad restringida. Así se tiene:

$$\Delta x_{reposo} = \frac{\Delta x_{movimiento}}{\sqrt{1-\frac{V^2}{c^2}}} \qquad (36\text{-}12)$$

Por otra parte, si medimos la masa de un objeto en movimiento, encontraremos, respecto de su masa medida en reposo, la relación:

$$m_{movimiento} = \frac{m_{reposo}}{\sqrt{1-\frac{V^2}{c^2}}} \qquad (36\text{-}13)$$

36.4 ¿Por qué no se puede sobrepasar la velocidad de la luz en el vacío?

En realidad esta afirmación es falsa. Nada prohíbe que dos sistemas se muevan a velocidad relativa V > c. De hecho, en la Cosmología actual se supone la existencia de galaxias que se alejan de la nuestra a velocidad mayor que la de la luz en el vacío. Lo que sí podemos asegurar es que, si dos sistemas se mueven con velocidad relativa mayor que la velocidad de la luz en el vacío, no podrán intercambiar información, serán, por tanto, universos separados. Y ello es así porque, en caso contrario, se violaría la

causalidad, que tanto la Relatividad Especial como la Relatividad General respetan escrupulosamente.

También podemos asegurar que dentro de un sistema de referencia no podemos acelerar un objeto para que llegue a alcanzar la velocidad de la luz. Si observamos la expresión (36-13), vemos que, cuando la velocidad del objeto se acerca a la velocidad de la luz, la masa se hace cada vez más grande, hasta tender al infinito cuando V se aproxima a c. Ello significa –no lo vamos a demostrar– que la energía que damos al objeto para acelerarlo, a medida que su velocidad se hace más grande, se invierte en aumentar su masa y no su velocidad. Este fenómeno está en el origen de la ecuación $E= m c^2$, que nos dice que la masa no es sino una forma de energía, y podemos obtener masa de la energía y viceversa.

No obstante, hay más razones. Examinemos la causalidad entre dos sistemas que se mueven a velocidad relativa $V > c$. Si el SRI2 se acerca al observador O1 a velocidad V superlumínica, y emite, en un instante, t_1, un rayo de luz codificada con la información sobre un suceso cualquiera, el sistema dejará atrás a la luz. Suponemos que en un instante de tiempo posterior, t_2, emite otro rayo de luz, con la información relativa al suceso consecuencia del anterior, es evidente que esta información llegará a O1 antes que la información emitida previamente por SRI2. En otras palabras, a O1 llegaría la información del efecto antes de llegar la información de la causa, violando el principio de causalidad. Hemos supuesto que SRI2 se acerca a SRI1 pero, ¿qué pasa si se están alejando? Volvamos al ejemplo de la plataforma. Si examinamos la expresión (36-7), obtendríamos un tiempo negativo de llegada de la luz al sensor, es decir, ¡nunca llegaría!, o equivalentemente, no tendríamos información del suceso del levantamiento de la banderola de la derecha de la plataforma.

En resumen, podemos afirmar que si dos SRI se mueven relativamente a velocidad superlumínica, nunca podrán intercambiar información y, a efectos de las fuerzas entre objetos, nunca podrán interaccionar.

La Teoría de la Relatividad Especial o Restringida, está hoy suficientemente contrastada por los experimentos y es aceptada universalmente por la comunidad científica, habiendo dado lugar a numerosas aplicaciones, entre las que destaca la conversión de materia en energía mediante la fisión y la fusión nucleares. Una de sus consecuencias más importantes es la no independencia del espacio y el tiempo que quedan, ahora, íntimamente ligados entre sí y al sistema de referencia de la observación.

Notas:
En este enlace el lector puede encontrar un vídeo, sacado del curso audiovisual de Física del Caltech de nombre “El Universo Mecánico y más allá”:
https://www.youtube.com/watch?v=ts8BV_fxLcQ

Julio Gutiérrez Muñoz
Doctor en Física
Catedrático de Universidad de Física Atómica, Molecular y Nuclear, jubilado.

Capítulo 37
¿En qué consiste la Teoría de la Relatividad General? ¿Por qué "pesa" la luz?

(Por Julio Gutiérrez Muñoz)

37.1 Introducción

En el capítulo anterior, epígrafe 36.2, hemos analizado el Principio de Relatividad, el cual establece que el estado de movimiento de un objeto no es absoluto y depende del estado de movimiento del observador, es decir, la velocidad de un objeto es relativa al Sistema de Referencia (SR); si alguien camina por un vagón de un tren en marcha, la velocidad medida por otro pasajero no es la misma que mediría un observador situado en reposo respecto de la vía. En el escenario de la Relatividad Especial, las aceleraciones y, en consecuencia, las fuerzas tampoco son invariantes, es decir, dependen también del SR; aspecto que no hemos tratado por salirse del alcance de este libro. Sin embargo, en un escenario de velocidades pequeñas frente a la velocidad de la luz, los cambios en las fuerzas y aceleraciones, al cambiar de SR, siempre que manejemos sistemas de referencia inerciales (no acelerados, SRI), son despreciables y se puede utilizar la Relatividad de Galileo. En caso de utilizar sistemas no inerciales (SRNI), el escenario se complica enormemente. Observado el movimiento de un objeto a velocidad constante en módulo y dirección en un SRI, en el seno de un SRNI puede verse animado de fuertes aceleraciones, lo cual es contradictorio con los valores nulos que mostrarían los acelerómetros solidarios con el objeto. La Teoría de la Relatividad General consiste en la extensión del Principio de Relatividad a los sistemas de referencia no inerciales, haciendo uso de la gravedad.

37.2 Fuerzas ficticias o de inercia

Para poder explicar tales movimientos acelerados aparentes en los SRNI aplicando las leyes de Newton, la única solución consiste en introducir unas fuerzas ficticias o de inercia: ***las llamadas fuerzas de inercia son fuerzas ficticias o aparentes que un observador no inercial percibe como reales, no teniendo existencia real en un sistema inercial. Dado su carácter ficticio, no pueden provocar cambios en el estado de movimiento de un objeto observado desde un sistema de referencia inercial***.

Estas fuerzas son muy diferentes según que el SRNI esté acelerado solo en el módulo de su velocidad, en cuyo caso el tratamiento matemático es muy simple, o su velocidad cambie de dirección, incluso aunque no cambie de módulo, apareciendo aceleraciones aparentes complejas, como es el caso de la aceleración de Coriolis que trataremos en el capítulo 70. Veamos algunos ejemplos.

Primero regresemos al tren y supongamos que un pasajero se encuentra en el centro del vagón, grabando la escena con una cámara de vídeo, sobre un monopatín que

carece de rozamiento con el suelo. Un observador acompañante está sentado y también graba las imágenes, lo mismo que hace un curioso situado junto a la vía. Si el tren frena, ¿qué ven y sienten nuestros tres protagonistas? El pasajero del monopatín no detecta aceleración alguna, pero ve acercarse la parte delantera del vagón aceleradamente. El pasajero que va sentado siente la fuerza de frenado, transmitida por el rozamiento con el asiento, pero observa que el monopatín se acelera con una "aparente fuerza" que tiene sentido contrario a la de frenado; es como si el pasajero del monopatín hubiera encendido un motor secreto. El espectador de la vía, evidentemente, no sentirá fuerza alguna, verá como el tren se detiene y al pasajero del monopatín seguir con la misma velocidad que llevaba antes del frenado del tren. Justamente lo contrario sucederá si el tren acelera, siendo ahora la parte trasera del vagón la que se acerca al monopatín.

Las imágenes grabadas dentro del tren, vistas por personas ajenas al experimento que no saben que se ha producido un cambio en la velocidad del vagón convirtiendo a este en un SRNI, necesitan introducir una "fuerza ficticia" para explicar el movimiento del viajero del monopatín y poder analizarlo utilizando la expresión $\boldsymbol{F} = m\,\boldsymbol{a}$. Las imágenes grabadas desde la vía no dan lugar a "engaño" alguno, están tomadas desde un SRI y muestran los movimientos en toda su pureza.

Pongamos un segundo ejemplo muy esclarecedor y que da lugar a no pocos malentendidos. Los astronautas de la Base Espacial Orbital nos envían sus imágenes, en las cuales los vemos flotando como si no estuvieran sujetos a la gravedad. Para explicar esa "ingravidez", en el SRNI que es la Base, se debe introducir una fuerza ficticia de sentido contrario a la de la gravedad –esta última está realmente presente y los mantiene en órbita–; esa "nueva fuerza" recibe el nombre de **centrífuga**, en compensación a la de la gravedad que está dirigida hacia el centro de la Tierra, denominada **centrípeta**. Visto por un observador inercial, los astronautas y la Base se encuentran en continua caída libre, sujetos a la aceleración de la gravedad, pero, como están animados de una velocidad perpendicular a la dirección de la fuerza gravitatoria, esta cambia continuamente la dirección de la velocidad del conjunto sin cambiar su módulo, y el sistema da vueltas alrededor de la Tierra a una distancia h de su centro dada por la expresión:

$$h = v^2 / g$$

siendo v la velocidad de la Base en relación al centro terrestre y g el valor de la aceleración de la gravedad a esa distancia.

De lo anterior se desprende que, cuando se hacen observaciones en sistemas no inerciales, debemos ser cuidadosos en distinguir qué aceleraciones son ficticias y cuáles son reales. Estas últimas se deben habitualmente a las ligaduras de los objetos con el sistema. Así, en el ejemplo del tren anterior hacíamos notar que el pasajero sentado era objeto de la aceleración real del tren transmitida por su ligadura al asiento. No obstante, ¿podemos realmente distinguir las fuerzas ficticias de las reales?

37.3 Principio de Equivalencia

Podemos también preguntarnos si las aceleraciones reales o ficticias son invariantes al cambiar de SR o relativas a cada sistema. En la relatividad de Galileo, en la cual las interacciones viajan a velocidad infinita, la respuesta es afirmativa para sistemas inerciales, pero no ocurre así en la relatividad einsteniana, donde aparecen términos adicionales al cambiar de SR, aunque el cambio se haga entre sistemas inerciales, que dan lugar a aceleraciones que llegaron a ser confundidas con una nueva interacción (caso de las fuerzas magnéticas). Además, queda por dilucidar el dilema de siempre: un sistema que está acelerado ¿lo está respecto de qué o de quién? Imaginemos nuestro laboratorio móvil en una nave espacial en medio de ninguna parte, ¿cómo podemos saber si estamos quietos o en movimiento y si éste es constante o acelerado?

Afortunadamente no somos capaces de saber si nos movemos con velocidad constante en módulo y dirección o estamos en reposo; una velocidad sin cambios no se distingue de otra. Esta propiedad dinámica no es sino la expresión experimental de relatividad de la velocidad y sería absurdo que se pudieran distinguir. Sin embargo, tan solo hace poco más de un siglo que, tras la invención del automóvil y la aviación, con ingenios capaces de desarrollar velocidades hasta entonces inalcanzadas, se predijo que a esas nuevas grandes velocidades (60 km/h) ocurrirían verdaderas catástrofes biológicas e incluso desintegraciones de la materia. Si bien no parecía haber diferencia al moverse a velocidades diferentes, ello era debido a lo bajas que eran las velocidades capaces de desarrollar, argumentaban.

En resumen, una velocidad sin cambios no se distingue de otra. Con un mar en calma, podemos realizar a bordo de un barco el mismo tipo de actividades que en tierra firme (jugar a la pelota, pasear por cubierta, tomar café en el bar sin verter el líquido, etc.); no notaremos diferencia alguna. Somos insensibles a nuestro movimiento de 30 000 m/s alrededor del Sol. Mas el Sol, a su vez, se mueve respecto del centro de la Vía Láctea, entonces, ¿cuál es nuestra verdadera velocidad de traslación? Sin embargo, no ocurre así con las aceleraciones. No solo podemos fabricar aparatos capaces de detectarlas aunque estemos en reposo en nuestro SR, sino que nuestro propio oído está diseñado para decirnos si nos movemos solidariamente con un SR acelerado. Es más, las aceleraciones grandes sí que producen esas catástrofes predichas para las altas velocidades.

Todo lo anterior es verdad incluso en el marco de la relatividad especial, pero Einstein fue más allá y se hizo una pregunta que él mismo consideró "*la idea más feliz de mi vida*": si un observador se encuentra en caída libre en un campo gravitatorio, ¿será capaz de detectar la aceleración que ha provocado la caída? En el ejemplo anterior de la Base Espacial Orbital parece que no es posible y, si los astronautas no tienen más referencias que las de su sistema, no serán capaces de distinguir si se encuentran en caída libre o en un lugar del Universo donde las fuerzas gravitatorias se han compensado completamente, porque en ningún caso sentirán su peso.

Pero aún hay más. También podemos preguntarnos si hay alguna forma de distinguir la gravedad de una aceleración debida a una fuerza no gravitatoria aplicada a un SRNI. Veamos unos experimentos mentales:

Experimento 1: Siguiendo la feliz idea de Einstein, supongamos un ascensor en caída libre. Un observador en su interior que deje caer a ambos lados dos objetos (tanto da

que sean iguales que no, pues la aceleración es la misma para ambos), no distinguirá esta situación de la encontrarse en un lugar del Universo donde las fuerzas gravitatorias se han compensado completamente y los objetos permanecerán sin caer, respecto del observador en ambas situaciones, en el punto en que se abandonaron.

Experimento 2: Supongamos un viajero del espacio que despierta de su hibernación tras un largo viaje y detecta estar sometido a una fuerza que lo acelera contra el suelo de la nave, de valor $\boldsymbol{a}$. Si no le es posible observar el exterior, no tendrá forma de decidir si está posado en un planeta cuya gravedad tiene valor $\boldsymbol{a}$, o bien se encuentra todavía en el espacio y los motores aceleran la nave con valor $\boldsymbol{a}$. En los dos casos, si abandona a ambos lados dos objetos, caerán hacia el suelo con aceleración $\boldsymbol{a}$ igualmente.

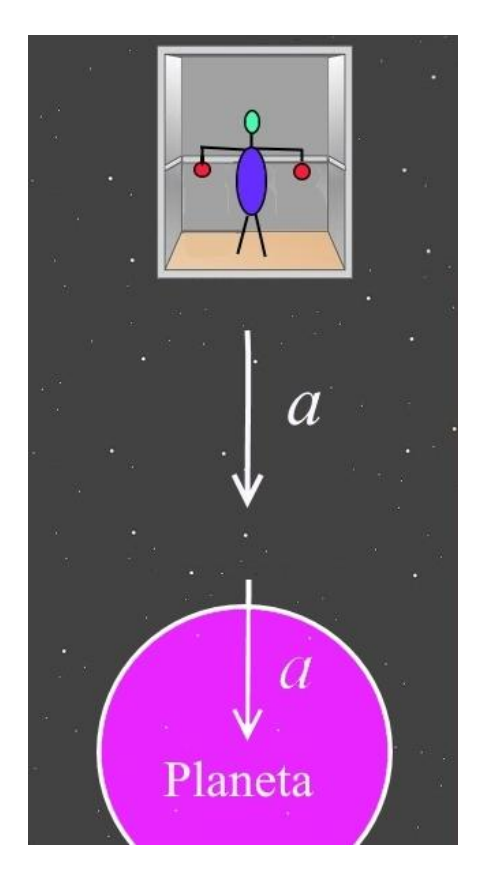
En analogía con el Principio de Relatividad, se enuncia un nuevo principio que, en este caso, se denomina **Principio de Equivalencia**: ***Un observador en caída libre en un campo gravitatorio uniforme y constante es equivalente a un observador inercial en ausencia de gravedad. Además, un observador en movimiento uniformemente acelerado es equivalente a un observador inercial en un campo gravitatorio uniforme y constante. Es imposible determinar la diferencia entre las situaciones anteriores a base de experimentos físicos.***

37.4 Generalización del principio de equivalencia

Seguramente el lector se habrá preguntado si realmente no hay posibilidad de diseñar un experimento físico para distinguir entre las opciones del enunciado anterior.

Para empezar, podría argumentar que, hasta hoy, nadie ha podido demostrar que la masa gravitatoria –la correspondiente a la Ley de Gravitación de Newton y responsable del peso de los objetos– y la masa inercial –la que aparece en la Segunda Ley de Newton ($\boldsymbol{F} = m\,\boldsymbol{a}$) responsable de la aceleración en repuesta a una fuerza de naturaleza cualquiera– sean iguales, aunque las tomamos como tales. Hay experimentos que demuestran que tienen el mismo valor con una gran precisión, pero no son estrictamente válidos. Si ambas masas fueran en realidad proporcionales aunque distintas, nos proporcionarían el mismo valor por nuestra forma de medirlas; bastaría cambiar el valor de la Constante de Gravitación Universal, G, para hacerlas diferentes pero proporcionales. Lo que se hace es asumir un **principio de equivalencia entre las masas gravitatoria e inercial** para zanjar la cuestión. El problema, sin embargo, no está resuelto y desde 2010 se realizan experimentos para demostrar la no equivalencia a nivel cuántico.

No obstante lo anterior, sí que es posible distinguir si el sistema está acelerado o sometido a la gravedad. Analicemos los experimentos 1 y 2 del epígrafe anterior. Las observaciones de ambos observadores serían las allí descritas si el campo gravitatorio fuera estrictamente homogéneo, pero para ello se necesita una distribución de materia que está muy lejos de ser real. En la Naturaleza los campos gravitatorios son causados por la presencia de materia y cada distribución de materia da lugar a una configuración específica de campo gravitatorio, en general inhomogénea.



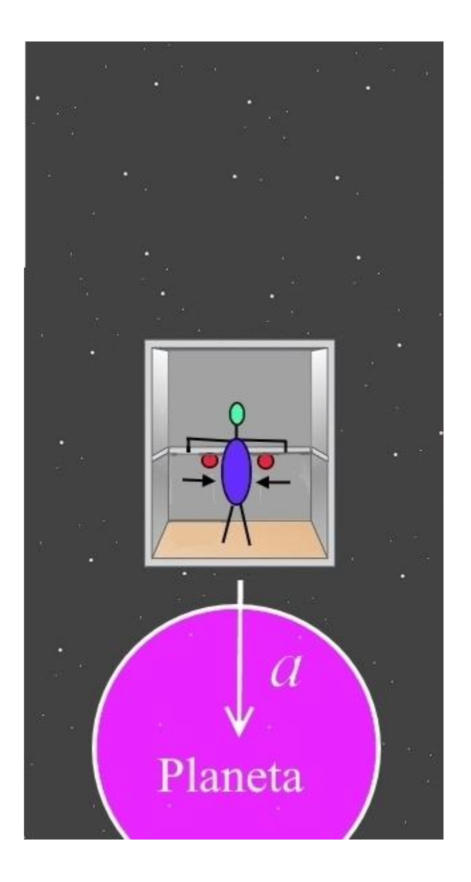


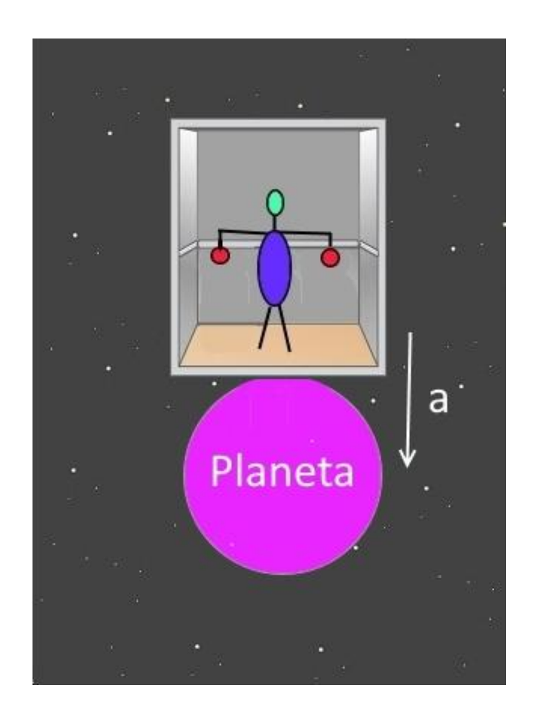
Figura 1: *Movimiento aparente de objetos abandonados en un sistema de referencia en caída libre en un campo gravitatorio.*

Experimento 3: Como muestra la figura 1, si la caída libre se produce en un planeta como la Tierra, observada desde un SRI, los objetos describen una trayectoria vertical desde su posición inicial que converge hacia el centro del planeta, por lo tanto, sus trayectorias de caída libre no son paralelas. La observación desde el SRNI mostraría a los objetos "flotar" a la altura a la que se abandonaron, pero los vería acercarse horizontalmente, sin razón aparente, lo cual no ocurre si las fuerzas gravitatorias han sido compensadas.

Experimento 4: Como muestran las figuras 2 y 3 la trayectoria de los objetos en su caída en el sistema en reposo sobre la superficie de un planeta son convergentes hacia su centro, mientras que en una situación de aceleración del sistema, las trayectorias serían paralelas.

Pero aún hay más, en un campo gravitatorio no homogéneo la gravedad disminuye con la distancia a la masa responsable de la fuerza, por consiguiente, si el sistema es muy grande "en altura", se producirá un efecto marea, es decir, las partes más cercanas a la masa estarán sometidas a una gravedad mayor que las alejadas.

En resumen, un campo gravitatorio constante es realmente una aproximación a pequeña escala, válida en una región restringida cerca de la superficie de objetos grandes, como hacemos habitualmente cuando nos movemos sobre la superficie terrestre.



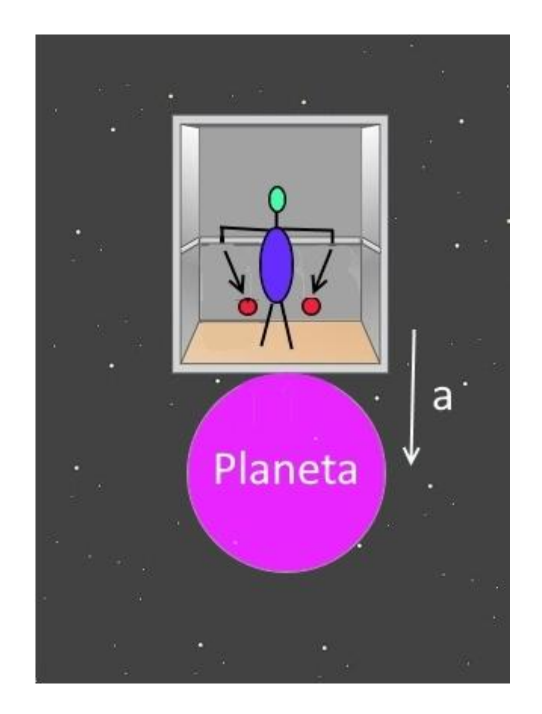
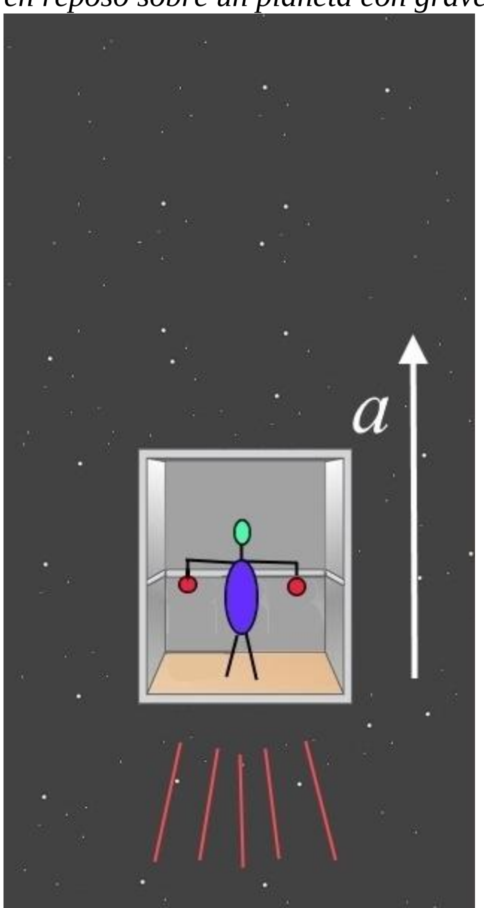


Figura 2: *Trayectoria de objetos abandonados en un sistema de referencia situado en reposo sobre un planeta con gravedad* ***a****.*



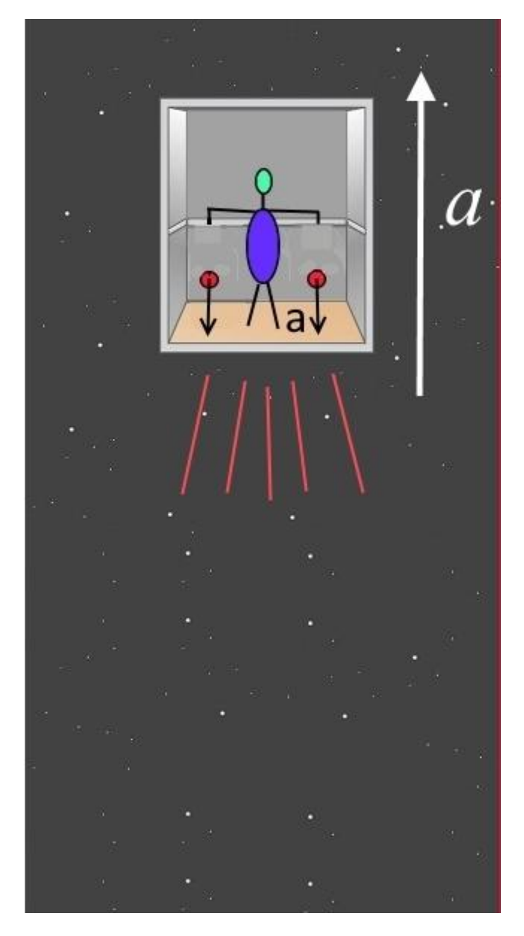


Figura 3: *Trayectoria de objetos abandonados en un sistema de referencia no inercial con aceleración* ***a****.*

Consecuentemente, si bien el Principio de Equivalencia fue históricamente importante para formular la teoría de la Relatividad General, solo es estrictamente

válido en experimentos de pequeñas dimensiones. Habrá que volver a enunciar el principio para hacerlo universal; puede quedar de esta forma: ***Un observador en caída libre en un campo gravitatorio no uniforme y constante es localmente equivalente a un observador inercial en ausencia de gravedad. Además, un observador en movimiento uniformemente acelerado es localmente equivalente a un observador inercial en un campo gravitatorio uniforme y constante. Es imposible determinar la diferencia entre las situaciones anteriores a base de experimentos físicos.***

La inclusión del término “localmente” implica que la Relatividad General solo será válida en regiones pequeñas del espacio, no pudiendo generalizar los resultados a sistemas muy grandes.

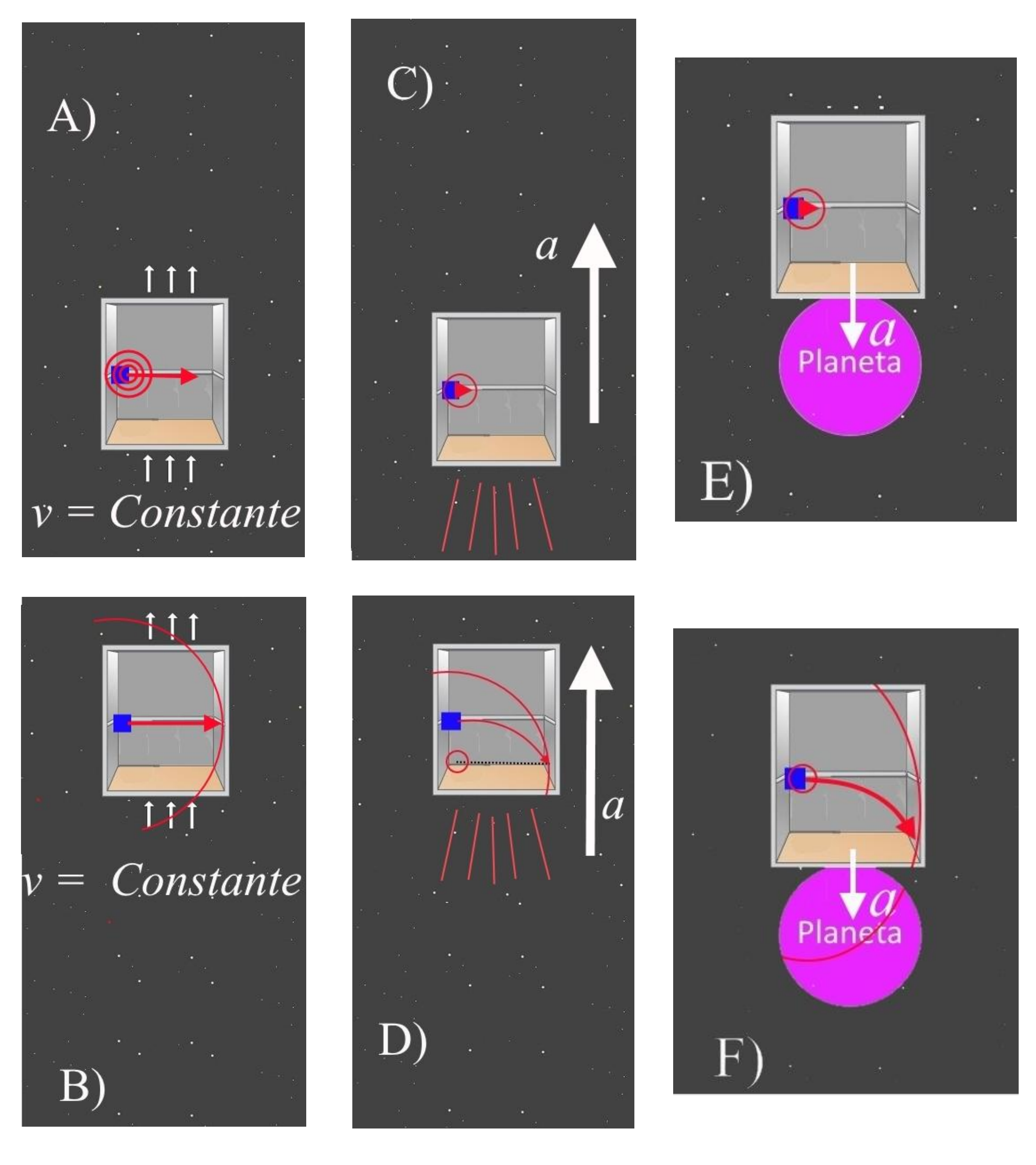


Figura 4: *En la fila superior se muestra la emisión de un rayo de luz en un instante de tiempo en tres sistemas de referencia, la fila inferior ilustra el punto en que se proyectará el spot luminoso en un instante posterior. Los SR son los siguientes:* A *y* B *SRI;* C *y* D *SRNI con aceleración* ***a****;* E *y* F *sistema en reposo en un planeta con gravedad* ***a****. Las curvas rojas indican las esferas o elipses de propagación de la luz.*

Una de las consecuencias más llamativas del Principio de Equivalencia consiste en la curvatura de la trayectoria de un rayo luminoso en presencia de un campo gravitatorio. Si localmente no podemos realizar ningún experimento físico para determinar nuestra situación en un sistema acelerado o en reposo en la superficie de un planeta, tampoco es válido un experimento con un rayo de luz. Así, en una nave espacial que no tenga referencias exteriores y que se mueva con aceleración ***a***, la luz deberá comportarse igual que si la nave se encuentra en reposo en un planeta de gravedad ***a***, siempre que se respete la condición de localidad. Si en un instante *t,* en el sistema acelerado, proyectamos horizontalmente un rayo de luz desde un lateral al de enfrente, las ondas han de estar centradas en el punto en que el foco emisor se encontraba en el instante inicial. Durante el tiempo que el rayo tarda en recorrer el camino, la pared de la nave se habrá movido aceleradamente y el rayo impactará más abajo, describiendo una trayectoria curva como muestra la figura 4D. Recordemos que la velocidad de la luz no puede estar afectada por el movimiento del emisor (véase el epígrafe 36.2). En el caso de la nave en reposo sobre un planeta, la luz deberá describir la misma trayectoria curva ilustrada por la figura 4F para cumplir el Principio de Equivalencia, en otras palabras, la luz "pesa" aunque carezca de masa. La primera comprobación empírica de la Teoría de la Relatividad General tuvo lugar durante los eclipses solares de 1919 y 1922, observando la curvatura de la luz –al pasar cerca del Sol– procedente de una estrella lejana.

37.5 Conclusiones

Como consecuencia de esa "atracción gravitatoria" de la luz, Einstein pensó que la gravedad podría no ser una fuerza fundamental de la Naturaleza y que el dilema se resolvería si conseguía encontrar las ecuaciones que dieran cuenta de los fenómenos gravitatorios como resultado de que el espacio y el tiempo –íntimamente ligados a partir de la relatividad especial (epígrafe 36.4) – se deformaban, curvándose, en presencia de una masa. Así, la luz, que debe propagarse por el espacio, seguiría la trayectoria que le marcara dicha curvatura. No obstante, la matemática de los espacios curvos es muy complicada y las consecuencias son difíciles de predecir, sobre todo porque apenas se han obtenido soluciones a las ecuaciones de Einstein. De entre ellas las más notorias son:

Explicación de la precesión anómala del perihelio del planeta Mercurio (confirmada la coincidencia entre datos reales y teóricos con una precisión de un 0,5%).

Posible existencia de agujeros negros (estos forman parte de los dogmas de fe de la Astrofísica, pero aún falta la comprobación experimental).

Atraso de los relojes en un campo gravitatorio (a tener en cuenta en los relojes de los satélites del GPS, fenómeno combinado con la dilatación temporal de la relatividad especial, véase la ecuación 36-11 en el epígrafe 36.3).

Alargamiento de las dimensiones de los objetos en la dirección de la aceleración de la gravedad (sin comprobar).

Disminución de la frecuencia de la luz cuando se aleja de un objeto masivo (comprobado en 1959 por Pound y Rebka).

Cambio de la orientación de un giróscopo en rotación en presencia de un campo gravitatorio (comprobado en mayo de 2011 por el satélite *Gravity Probe B*).

Existencia de las ondas gravitatorias (posible comprobación experimental en este mismo año 2016).

A pesar de su aparente éxito, la Teoría de la Relatividad General, no es universalmente aceptada todavía. Hay muchos aspectos oscuros, varias contradicciones y algunos científicos siguen manteniendo la idea de que la gravedad es una fuerza fundamental y no una consecuencia de la curvatura del espacio-tiempo. De hecho, las ondas gravitacionales no son solo un resultado de las ecuaciones de Einstein, también se obtienen, mediante unos sencillos cálculos, al aplicar al campo gravitatorio las trasformaciones de Lorentz, las cuales proporcionan el cambio de sistema de referencia en la relatividad especial. Entre las contradicciones destacan dos por su evidencia.

En primer lugar, los agujeros negros son una consecuencia matemática de las ecuaciones de Einstein, mas, en sus proximidades, la localidad, elemento esencial del Principio de Equivalencia, se viola. Efectivamente, dada la intensidad de la gravitación en las proximidades de uno de estos agujeros: 1) el efecto marea sería detectable a pequeñas distancias; y 2) como el tamaño de estas singularidades es muy pequeño, los efectos ilustrados por las figuras 1 y 2 serían muy acusados. Entonces, si el Principio de Equivalencia no es válido, las ecuaciones de Einstein no son aplicables en ese entorno y sus predicciones serían falsas.

En segundo lugar, si debido al límite de velocidad impuesto por la relatividad especial, ninguna información puede escapar de un agujero negro, ¿cómo se entiende que la masa almacenada en uno de ellos pueda ser detectada por los cuerpos que se encuentran a cualquier distancia? Los puristas dicen que, al no ser una fuerza debida a una interacción, sino a la deformación del espacio-tiempo, no existe tal problema; pero la información no deja de ser información.

Julio Gutiérrez Muñoz
Doctor en Física
Catedrático de Universidad de Física Atómica, Molecular y Nuclear, jubilado.

Capítulo 38
¿Por qué la luz se mueve a la velocidad de la luz?
(Por Alberto Aparici)

Seguro que os parece que ésta es una pregunta un poco tonta.

Y si Albert Einstein no hubiera descubierto la teoría de la relatividad como lo hizo, probablemente lo sería. Pero pensad lo siguiente: la relatividad cambió lo que pensábamos sobre el espacio y el tiempo. Nos descubrió que los relojes pueden marcar el tiempo a diferentes velocidades, y que las imágenes que vemos de objetos lejanos pueden aparecer deformadas debido a su movimiento o al campo gravitatorio. En los capítulos anteriores, sobre la teoría de la relatividad, se ha dado buena cuenta de muchos de estos efectos, y sabemos que la velocidad de la luz juega un papel importante en todos ellos: a velocidades bajas, como las de nuestra vida cotidiana, estos efectos son imperceptibles, pero se van haciendo más y más patentes cuando el movimiento es muy rápido, cercano a la velocidad de la luz. La pregunta es ¿por qué la luz? ¿Qué tiene la luz de tan especial para que tantos fenómenos físicos giren a su alrededor?

Para entenderlo, primero deberíamos recordar algo: a la luz le da igual lo rápido que nos movamos, ella siempre se mueve igual. Por extraño que parezca, éste es un hecho experimental muy bien establecido: la velocidad de la luz en el vacío es siempre 299.792,458 km/s. Si tuviéramos una nave espacial que pudiera moverse a 200.000 km/s y tratásemos de "perseguir" a un rayo de luz, veríamos que la luz huye de nosotros a 299.792,458 km/s. Si aceleráramos con nuestra nave espacial hasta 299.000 km/s el rayo de luz seguiría alejándose de nosotros a 299.792,458 km/s. Por desgracia no tenemos la tecnología para hacer estos experimentos con humanos en naves espaciales, pero sí podemos comprobarlo utilizando partículas subatómicas, que se mueven a velocidades altísimas, y mediante algunos dispositivos experimentales verdaderamente ingeniosos.

La física, desde luego, no se queda de brazos cruzados mientras la luz hace todas estas cosas extraordinarias: para que todos los observadores vean moverse a la luz a la misma velocidad las propias leyes del movimiento deben cambiar. Cosas aparentemente sencillas, como que si vamos en un coche a 50 km/h y el coche de delante va a 60 km/h nosotros lo veremos alejarse a 10 km/h, sabemos que no son estrictamente verdad. Son *casi* verdad cuando las velocidades son pequeñas, como 50 km/h o 10 km/h, pero dejan de ser verdad cuando nos acercamos al leviatán de las velocidades: la velocidad de la luz. Pero de nuevo ¿por qué la naturaleza ha escogido a la luz? ¿Por qué no a los electrones o a las ballenas o a cualquier otra cosa?

Un famoso experimento mental de Einstein nos da una pista para responder a esta pregunta. Einstein se preguntó qué sucedería si pudiésemos movernos a la misma velocidad que un rayo de luz, si pudiéramos ver a la luz en reposo, congelada delante de

nosotros. A lo largo del siglo XIX se había descubierto que la luz es una onda electromagnética, un pulso de energía transportado por campos eléctricos y magnéticos en movimiento. Normalmente la electricidad y el magnetismo van ligados a cargas eléctricas: allá donde hay un electrón éste genera un débil campo eléctrico, y si el electrón se pone en movimiento tendremos, además, un campo magnético. Pero también hay otra manera de producir estos campos: si un campo eléctrico cambia con el tiempo parte de su energía se va a emplear en generar un campo magnético; y lo contrario también es verdad: un campo magnético variable generará inmediatamente un campo eléctrico. Este fenómeno, por el cual la electricidad genera magnetismo y el magnetismo genera electricidad, se llama *inducción electromagnética*.

Una onda electromagnética consiste en un campo eléctrico variable que genera por inducción un campo magnético un poco por delante de él. Este campo magnético también es variable, y genera a su vez un campo eléctrico un poco por delante. De esta forma, el campo eléctrico y el magnético se van “turnando” y se mueven por el espacio como si tiraran el uno del otro, empujándose para dar el siguiente paso. Normalmente las ondas electromagnéticas tienen su origen en un conjunto de cargas que se mueven de la forma adecuada –por ejemplo, los electrones que se mueven dentro de una antena–, pero una vez son liberadas al espacio se propagan de manera autónoma y ya no necesitan a las cargas porque es la inducción la que las sostiene. No es la luz el único fenómeno que se podía explicar mediante ondas electromagnéticas: las ondas de radio, los infrarrojos, y después también los rayos ultravioleta o los rayos X, son diferentes tipos de ondas electromagnéticas que se diferencian en que unas transportan más energía que otras.

Einstein, pues, se planteó qué sucedería si pudiésemos perseguir a una de estas ondas, alcanzarla y verla congelada delante de nuestros ojos. La conclusión fue bastante desagradable: la onda debería desaparecer, volatilizarse en el momento en que la alcanzáramos. Como el fundamento de la inducción es que los campos que forman la onda estén cambiando continuamente, si logramos “pararlos” la inducción también se para y la onda, sencillamente, deja de existir. En física siempre es mala noticia que las cosas se esfumen, y en este caso la noticia no tenía ni pies ni cabeza: si Einstein estaba en lo cierto, un observador parado podría ver una onda electromagnética viajando, por ejemplo, de una antena a otra entre dos edificios, y transportando información entre ambos lugares. Sin embargo, el observador que persigue a la onda y logra que desaparezca no vería nada viajando de un edificio a otro: la información le llegaría al receptor sin nada que la transporte, como si fuera magia. Einstein concluyó que el segundo observador no podía existir. La física, fuera como fuese, tenía que ingeniárselas para que nadie pudiese alcanzar a un rayo de luz o a una onda de radio, o de lo contrario éstas podrían aparecer y desaparecer como un fantasma, y con ellas la energía que transportan.

Esta idea fue la que llevó a Einstein a proponer que tal vez la velocidad de la luz no se viera afectada por el movimiento del observador. Era una buena manera de hacer a la luz “inalcanzable”, porque por muy rápido que nos moviéramos el rayo de luz siempre se estaba alejando de nosotros. En el fondo, el motivo no era que la luz fuese muy especial, sino que la luz, por su propia naturaleza, tiene que estar en movimiento o de lo contrario desaparecería. Lo que Einstein descubrió cuando dijo que la velocidad de la luz es constante es que hay una especie de “velocidad de las cosas que son solo movimiento”, y que esa velocidad tenía que ser la misma para todos los observadores en todos los lugares del universo.

Hay otra forma de ver este mismo fenómeno. Es archifamosa la fórmula de la relatividad especial $E=mc^2$, que expresa que la masa es una forma de energía y que la energía se puede transformar en masa. Es menos conocido que ésta es una fórmula para objetos en reposo, y su forma completa es $E=\square mc^2$, donde $\square$ es el *factor de Lorentz*, una cantidad que depende de la velocidad del objeto. El factor de Lorentz vale 1 cuando la velocidad es cero, y se hace más y más grande, hasta infinito, a medida que la velocidad aumenta. De esta manera la fórmula nos informa de que un objeto va acumulando más y más energía a medida que su velocidad aumenta, pero que incluso cuando su velocidad es cero posee algo de energía: la que está almacenada en su masa.

Esta fórmula es válida para casi todos los objetos de nuestro universo. La luz, y las ondas electromagnéticas en general, es uno de los que *no* cumplen esta relación. La razón es sencilla: la luz no almacena ninguna energía a velocidad cero. Como Einstein intuyó, si un rayo de luz se para su energía se volatiliza. Los físicos solemos expresar esto diciendo que la *masa* de la luz es cero, $m_{luz}c^2 = E_{luz}(v=0) = 0$, o lo que es lo mismo: que la luz almacena toda su energía en forma de movimiento.

No es la única que lo hace. Como hemos dicho antes, todas las ondas electromagnéticas comparten esa propiedad: ondas de radio, infrarrojos, rayos ultravioleta, rayos X, rayos gamma… También se comportan así las *ondas gravitacionales*, las ondulaciones en el espacio-tiempo que transportan la gravedad de un lugar a otro. Y se ha propuesto la existencia de otras partículas que deberían tener masa cero y moverse a la velocidad de la luz, aunque ninguna de ellas ha sido descubierta, al menos por ahora.

En definitiva, ahora podemos dar respuesta a nuestra pregunta: el universo no tiene ninguna predilección especial por la luz, pero las leyes de la física reconocen la existencia de una serie de objetos que, por una razón o por otra, solo pueden existir si están en movimiento. Para esos objetos la física reserva una velocidad especial, c=299.792,458 km/s, que es la misma en todo el universo y para todos los observadores y que, además, no puede ser alcanzada por ningún observador, de forma que estos objetos especiales aparezcan en movimiento a los ojos de todo el mundo. Nosotros llamamos a esa c “velocidad de la luz” únicamente porque la luz fue el primero de esos objetos que pudimos observar, y el que mejor conocíamos en el momento en que la física estuvo madura para abordar esta cuestión. Pero igualmente podríamos decir que c es “la velocidad del movimiento absoluto”, y que la relatividad, irónicamente, no deja de ser una teoría sobre qué le pasa a la física cuando entra en escena un movimiento que no es relativo, sino absoluto.

Alberto Aparici
Doctor en Física
Instituto de Física Corpuscular, Universidad de Valencia, Onda Cero

Capítulo 39
El efecto Doppler o “corrimiento hacia el rojo” (redshift) de una onda electromagnética, ¿es posible saber si se debe a la velocidad del objeto o a la acción de un campo gravitatorio? ¿Y qué es el redshift cosmológico?
(Por Paola Marziani)

El efecto Doppler es un fenómeno físico muy básico que ocurre cuando un objeto que emite ondas, como por ejemplo la luz (ondas electromagnéticas) o el sonido (ondas acústicas), está en movimiento con respecto a un receptor (observador u oyente en el caso de la luz o sonido). Cuando las fuentes de emisión se están alejando del observador, las ondas sucesivas se emiten desde una posición más lejana al observador. Por lo tanto, el observador recibe cada onda después de un tiempo más largo respecto al que recibiría en el caso en que la fuente estuviera en reposo; por este motivo, la distancia temporal entre las crestas de la onda se incrementa y la frecuencia se reduce. Lo contrario sucede cuando la fuente de ondas se aproxima al observador. Por consiguiente, el efecto Doppler puede dar lugar a un cambio de longitud de onda que puede ser positivo si la fuente se está alejando, pero también negativo si se está acercando.

El efecto Doppler clásico relaciona la velocidad de la fuente con la velocidad de propagación de la señal, es decir, v / c, donde c es la velocidad de la luz o la velocidad del sonido en el medio en el que la señal se propaga (que podría ser por ejemplo el vacío o aire para la luz y el aire o el agua para el sonido). El corrimiento de longitud de onda es simplemente proporcional a la relación $\Delta\lambda/ \lambda_0 = (\lambda-\lambda_0)/\lambda_0 = v / c$, donde λ_0 es la longitud de onda de la luz emitida por una fuente en reposo y puede tomar un signo negativo o positivo si la fuente se acerca o se aleja.

Una fuente que se aleja dará siempre un aumento de la longitud de onda. Con referencia a la longitud de onda de la luz visible, se dice a menudo un corrimiento al rojo, es decir, un desplazamiento hacia el rojo. Este término, en astronomía se utiliza para cada frecuencia, incluso el dominio de rayos X y de radio para indicar un aumento de la longitud de onda.

La fórmula anterior es una de las más importantes en todo el campo de la física. Las cosas empiezan a ser menos sencillas si tenemos en cuenta a una fuente que se está moviendo a una fracción no despreciable de la velocidad de la luz. En este caso habrá un término adicional que siempre va a dar un corrimiento al rojo, independiente de la dirección del movimiento. Este término es un efecto puro de la relatividad especial, en relación con la incapacidad de sincronizar nuestros relojes con el de la fuente de movimiento y puede ser arbitrariamente grande cuando la velocidad de la fuente se acerque a la velocidad de la luz.

El corrimiento al rojo gravitacional tiene un origen independiente al efecto Doppler. No solo que siempre se corre al rojo, sino también necesita un campo gravitacional muy fuerte para ser detectado en fuentes astrofísicas.

Es interesante tener en cuenta que, tan pronto como se descubrieron a los quásares, algunos astrónomos pensaban que el corrimiento al rojo no podía ser debido a “la velocidad de recesión de las galaxias” (es decir, a la distancia en una formulación obsoleta y incorrecta), sino que tenía que ser de origen gravitacional. Esta idea fue rápidamente abandonada debido a paradojas que plantean la necesidad de estar muy cerca de una masa por obtener el gran corrimiento al rojo de los quásares.

El pequeño tamaño de la fuente ponía un límite muy fuerte a la cantidad de radiación que podría producirse. Y de hecho es necesario estar muy cerca por tener un corrimiento al rojo no despreciable debido a que el desplazamiento al rojo gravitacional es proporcional al cociente entre la masa y la distancia de un objeto (no es que la masa tiene que ser grande, sino es un asunto del cociente masa-distancia).

El corrimiento al rojo gravitacional puede aumentar indefinidamente en una fuente que se acerca de un objeto masivo. En un contexto astrofísico, los casos más extremos son los agujeros negros y las estrellas de neutrones. En el caso de un agujero negro, una fuente vista por un observador distante nunca alcanzará el horizonte de eventos del agujero negro, pero el observador verá la luz emitida desplazada a una longitud de onda más larga.

Si originalmente la fuente estaba emitiendo luz visible, y si fuera posible seguir la luz desde la fuente hacia su caída al agujero negro en todas las frecuencias, el receptor detectará las ondas en el óptico y luego en el IR cercano, a continuación, en el IR lejano, en el sub-mm, y luego a más y mayores longitudes de onda en el dominio del radio. La luz emitida en las proximidades de un agujero negro aparecerá desplazada hacia el rojo debido a la pérdida de energía para superar el profundo potencial gravitacional del agujero negro.

El campo gravitacional puede crear un corrimiento al rojo arbitrariamente grande. ¿Qué pasa con el corrimiento al rojo de las fuentes astronómicas a gran distancia, como las galaxias y quásares? Como se ha mencionado, muchos años atrás la gente hablaba de velocidad de recesión de las galaxias. Edwin Hubble en el 1925 encontró una relación directa entre el corrimiento al rojo y la distancia de las mismas galaxias, es decir $v = c\ z = H0\ d$, donde $H0$ es la constante de Hubble estimada alrededor de 70 km / s / Mpc.

Actualmente, esta ley sigue siendo válida y es utilizada por los astrónomos si $z << 1$. Sin embargo, hoy en día existen cientos de miles de galaxias y quásares conocidos con $z >> 1$, incluso hasta $z = 7$. Puesto que somos capaces de medir la velocidad solo a lo largo de nuestra línea de visión (la velocidad radial) para los objetos distantes, el corrimiento al rojo no se puede asociar a una velocidad física a menos que se cumplan condiciones poco realistas y muy especiales.

El corrimiento al rojo de galaxias y quásares se explica en el contexto de un Universo en expansión, al que se llama propiamente expansión cosmológica del Universo, donde el tejido del espacio se está expandiendo después de un estado denso y caliente que en su origen fue una singularidad (no se trata de una expansión en un volumen previamente existente, es el mismo espacio-tiempo que se expande).

El desplazamiento hacia el rojo vuelve a ser no solo un marcador de la distancia, sino también un indicador de la época cósmica. Y hay una diferencia física fundamental con un desplazamiento al rojo debido al efecto Doppler: las fuentes no se mueven una con respecto a la otra, sino están retrocediendo solo porque el espacio-tiempo se está expandiendo. Nuestras reglas son más grandes en espesor y los relojes están corriendo más lentamente con respecto a los de las fuentes distantes. A una distancia definida por un corrimiento al rojo z, los relojes estarían corriendo más rápido por un factor (1 + z).

Se agradece la ayuda de las doctoras **Alenka Negrete** y **Alba Grieco** por una revisión precisa de este capítulo.

Paola Marziani
Doctora en Astrofísica
Investigadora / INAF - Osservatorio Astronomico di Padova

Capítulo 40
¿Qué es la teoría cuántica?

(Por Gerardo Ortiz)

La teoría cuántica fue inicialmente concebida como un intento de explicar el espectro de energía de un cuerpo negro, un objeto físico idealizado que absorbe toda la radiación incidente antes de volver a emitirla. Según la física clásica, a una temperatura fija, la energía emitida debería aumentar, sin límite, con la frecuencia de la onda radiante emitida. Sin embargo, los experimentos contradecían esta predicción clásica. Fue Max Planck quien resolvió la paradoja introduciendo el concepto de quanta en 1900. Esencialmente, Planck postuló que la radiación solo podía ser emitida en unidades discretas, o quanta, de energía. Esta suposición aparentemente simple pero radical implicaba que los modos de frecuencia más alta eran menos probables de ser excitados por lo que su energía promedio disminuiría con la frecuencia. Más tarde se observó que las líneas espectrales discretas de los átomos, que representan las transiciones entre los niveles estables de energía atómica, no podían explicarse por la teoría clásica del electromagnetismo, ya que los electrones (partículas cargadas negativamente que componen el átomo) caerían en espiral hacia el núcleo, haciendo que el átomo y, por lo tanto, toda la materia se volviese inestable. Fue Niels Bohr quien resolvió este rompecabezas en 1913 extendiendo las ideas iniciales de Planck sobre la cuantización a otras magnitudes dinámicas como el momento angular, la propiedad que caracteriza la inercia rotatoria de un objeto físico, obligando así a los electrones a moverse en órbitas particulares bien definidas.

Aunque estas ideas iniciales eran fundamentales, cada vez estaba más claro que algo profundo ocurría en el micro-mundo y aún no se había descubierto. Fue entonces el momento de dos científicos notables, Werner Heisenberg (1925) y Erwin Schrödinger (1926), que independientemente introdujeron formulaciones equivalentes de una nueva mecánica de ondas, más tarde conocida como teoría cuántica no relativista. Un aspecto revolucionario clave de esta nueva mecánica de ondas era que las condiciones externas al sistema y un aparato de medición podían determinar si un sistema físico se comporta como una partícula o una onda. Para capturar esta dualidad llamada onda-partícula, el sistema no podía ser simplemente especificado por su posición y velocidad, sino por un objeto matemático llamado función de onda, un catálogo de potencialidades. El carácter de onda de esta función de estado conduce naturalmente al fenómeno de la interferencia, algo observado normalmente en ondas ordinarias que se manifiestan en la luz o el agua.

Sin embargo, en la mecánica cuántica este fenómeno se manifiesta espectacularmente como resultado de la medición de un conjunto de partículas cuánticas igualmente preparadas que inciden, por ejemplo, en una pantalla con doble rendija. Lo que es más importante, esta manera particular de caracterizar el estado de un sistema implicaba una descripción probabilística de la naturaleza, una salida dramática de los sueños de Pierre-Simon Laplace (1814) de una descripción completamente determinista

de nuestra realidad externa. En consecuencia, la medición simultánea de dos propiedades complementarias del sistema, como la posición y el momento de un electrón, está sujeta a imprecisión; cuanto más precisamente tratamos de medir una de esas propiedades, menos precisa es la otra, algo conocido popularmente como el principio de incertidumbre.

En 1935, Albert Einstein, Boris Podolsky y Nathan Rosen sugirieron un experimento mental con consecuencias asombrosas. Se dieron cuenta de que un estado de un sistema cuántico compuesto no está necesariamente determinado por los estados de sus subsistemas constituyentes. Esta es una característica clave de los fenómenos cuánticos que más tarde Schrödinger denominó entrelazamiento, un recurso único que proporciona correlaciones no locales entre subsistemas que no admiten ninguna interpretación clásica local. Es esta no localidad lo que es difícil de comprender (y de hecho estaba en el centro de una serie de controversias que Einstein expresó con la teoría cuántica), pero en las últimas décadas los avances en la óptica moderna han hecho posible producir rutinariamente estados entrelazados de fotones (partículas de luz) usando láseres y las propiedades ópticas no lineales de ciertos cristales. Utilizando estas técnicas, muchos investigadores han podido confirmar las predicciones de la mecánica cuántica (y refutar teorías alternativas que involucran "variables ocultas" las cuales expresan la aleatoriedad cuántica en términos de elementos desconocidos de la aparente realidad externa).

El formalismo de la mecánica cuántica se ha extendido durante los últimos 80 años para acomodar la relatividad especial y la teoría de campos, y la comprensión de sus detalles ha sido esencial para el desarrollo de gran parte de la tecnología moderna (desde la comprensión mecanicista de la química sintética y física nuclear hasta las bases de los semiconductores y las industrias de almacenamiento magnético, por citar solo algunos ejemplos).

La teoría cuántica rechaza la visión clásica de la realidad física. Define un marco teórico fundamental que representa nuestra mejor descripción posible del mundo externo conocido, y junto con la teoría de la relatividad, forma la base de la física moderna. Se aplica a la descripción de los fenómenos naturales en todas las escalas, desde el micro hasta el macro-mundo, incluyendo los problemas fundamentales de la cosmología. A pesar de todos sus éxitos hay un aspecto en el formalismo que es difícil de tragar; en cierto sentido, los sistemas no poseen propiedades definidas (objetivas) hasta que medimos dichas propiedades, y la medida real siempre encuentra el sistema en un estado definido, no en la superposición de estados que conforman la función de onda real.

Uno solo puede predecir la probabilidad de un determinado resultado de una medición, pero no puede predecir el "colapso" real a un estado definido. Esto se conoce como la paradoja de la medición cuántica. Los dos principios más fundamentales que diferencian el realismo cuántico del realismo clásico son la no localidad genérica de las correlaciones cuánticas, o el entrelazamiento, y el hecho más general de que la mecánica cuántica es intrínsecamente contextual, lo que significa que el resultado de la medición de cualquier propiedad de un sistema depende de la configuración experimental específica que se utiliza para medir esa propiedad. En otras palabras, el resultado de una medición no debe considerarse como información revelada que simplemente está escondida de nosotros, es decir, preexistente e independiente del contexto de esta medición.

El entrelazamiento es ciertamente la propiedad que se encuentra en el corazón de la riqueza y rareza del mundo de la mecánica cuántica. Ahora nos damos cuenta de que estas correlaciones no locales están detrás de algunos de los problemas más desconcertantes de la física de los materiales, y también han llevado a revoluciones en la computación, comunicación (por ejemplo, permitir protocolos cuánticos de teletransporte), criptografía y otros campos. Recientemente, la promesa del entrelazamiento para abrir inusitados horizontes técnicos tanto como nuestra capacidad de controlar y comprender este aspecto singular del universo cuántico se ha descrito como la segunda revolución cuántica. Los llamados ordenadores cuánticos prometen proporcionar una aceleración exponencial en la resolución de ciertos problemas que fundamentalmente están más allá de las capacidades de los ordenadores convencionales, y se cree que el entrelazamiento es un ingrediente esencial para comprender y desbloquear el poder de la computación cuántica.

Se agradece a D. **Felipe José Ramos Calderón**, piloto de aviación, estudiante de Física (UNED) y buen amigo, por la traducción precisa de este capítulo.

Gerardo Ortiz
Doctor en Física
Professor of Condensed Matter Physics, Indiana University

Capítulo 41
¿Cuál es la función y la utilidad de los tan aclamados diagramas de Feynman?
(Por Bartolo Luque)

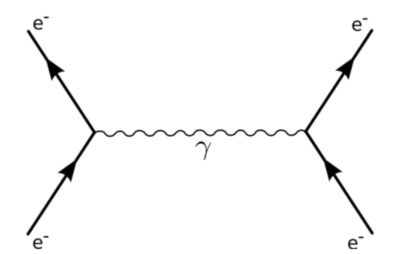
El Modelo Estándar, nacido en los años 70 del siglo pasado, es la descripción más contrastada que nos ofrece la física actualmente de la estructura fundamental de la materia y el vacío. El Modelo Estándar, que toma como base una serie de partículas elementales y cuatro interacciones fundamentales, es en realidad un conjunto de teorías. No todas ellas gozan de igual precisión, y en particular, "La joya de la corona", la teoría más precisa que disponemos, es la electrodinámica cuántica (QED, *Quantum Electrodynamics* en su acrónimo inglés). La QED es la teoría cuántico-relativista que describe cómo se producen las interacciones electromagnéticas entre partículas, y cómo interaccionan la luz y la materia. Las interacciones electromagnéticas son las fuerzas físicas que rigen, por ejemplo, cómo cargas del mismo signo se repelen entre sí y de signo opuesto, se atraen. En física clásica la descripción de estas interacciones vienen mediadas por el campo electromagnético. Pero en la QED, los electrones y demás partículas fundamentales intercambian fotones virtuales que hacen de portadores de la fuerza electromagnética. Una partícula virtual es una deuda transitoria de la naturaleza: aparece de la nada tomando prestada energía del vacío y, tras una breve vida, desaparece devolviendo la energía tomada en préstamo; siempre según las reglas establecidas por el principio de incertidumbre de Heisenberg.

El primer paso hacia la teoría lo dio en 1928 el gran físico Paul Dirac, al combinar la teoría cuántica con la relatividad especial para obtener una ecuación de ondas relativista para el electrón. Pero a pesar de los esfuerzos de muchos teóricos, la teoría presentaba un par de problemas de calado al describir el resultado de una interacción electromagnética entre partículas: uno era la aparición de cantidades infinitas en los cálculos y otro era el sinfín de términos que debían computarse, que convertían un problema en una pesadilla algebraica. Además, la teoría no daba cuenta de algunos desajustes entre medidas experimentales y resultados teóricos, como por ejemplo, la llamada constante de estructura fina que puede medirse a través del momento magnético anómalo del electrón. Así que en sus inicios la QED se mostraba confusa y refractaria al cálculo. Hasta que, de manera independiente, los físicos Richard P. Feynman, Julian Schwinger y Sin-Ichiro Tomonaga solventaron la cuestión; extremo que les valió el Premio Nobel de Física en 1965.

Schwinger y Tomonaga propusieron métodos matemáticos extremadamente complicados y poco prácticos de manejar. Mientras que Feynman consiguió presentar los cálculos de una manera intuitiva y solventar algunos de los molestos infinitos mediante una técnica llamada "renormalización". Para simplificar los pantagruélicos cálculos que aparecían en la QED Feynman propuso sus famosos diagramas, la base de su método. Fue Freeman Dyson quien demostró la equivalencia matemática de los tres métodos y ayudó a popularizar los diagramas de Feynman, que traspasaron enseguida el

área donde nacieron para aplicarse a la física nuclear o a la teoría del estado sólido. Los físicos teóricos, gracias a ellos, fueron capaces de realizar cálculos que en el momento resultaban inabordables. Del mismo modo que hoy la computación ha supuesto una revolución en muchas áreas de la ciencia, los diagramas de Feynman supusieron una revolución en la física.

Veamos con algo de detalle uno de los ejemplos más sencillos: la repulsión entre dos electrones. Clásicamente, para determinar la trayectoria de los electrones, usaríamos el campo eléctrico y consideraciones como la conservación de la energía. Pero la QED está gobernada por la cuántica de campos y una de las extraordinarias cosas que nos muestra esta teoría es que el campo eléctrico está cuantizado, constituido por fotones. La matemática de la física de partículas elementales es conocida como teoría cuántica de campos. Esta teoría describe como se propagan los campos cuánticos que describen las partículas, como colisionan e interaccionan entre sí, como se producen nuevas partículas, etc. Todos los procesos que se observan en los grandes aceleradores de partículas (como el LHC y el Tevatrón) se describen por esta teoría y la QED es su hija más perfecta. De modo que cuando dos electrones se acercan se pueden repeler, por ejemplo, a través de la emisión de un fotón por parte de uno de los electrones que será absorbido por el otro. Podemos dibujar fácilmente el diagrama de Feynman de este proceso:



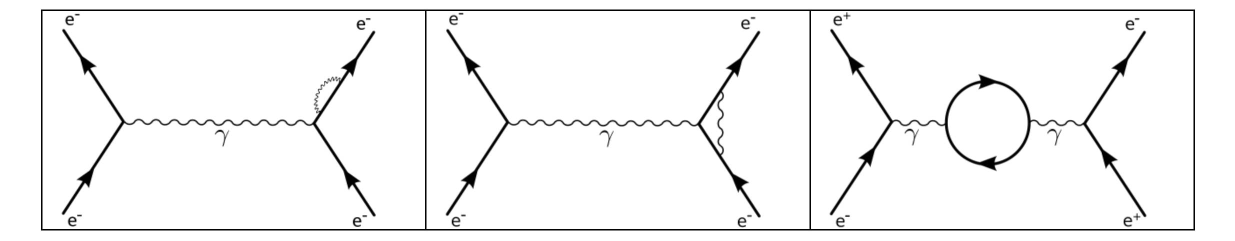
En todo diagrama de Feynman, las líneas rectas con flechas representan partículas de materia. En nuestro caso tenemos dos electrones e^-. El tiempo en nuestro diagrama transcurre de abajo hacia arriba y el eje horizontal puede interpretarse como el espacio (es usual encontrarnos los diagramas con el tiempo y el espacio intercambiados respecto al ejemplo que presentamos). Así, en la parte inferior del diagrama, tenemos las "partículas incidentes", que en nuestro caso son dos electrones que colisionarán entre sí; y en la parte superior tenemos las "partículas resultantes" después de la interacción, que, en este caso, siguen siendo nuestros dos electrones. El diagrama es un esquema taquigráfico y las líneas no deben entenderse como trayectorias de las partículas. Los vértices son aquellos puntos del diagrama donde concurren, por lo menos, tres segmentos. En ellos se producen las interacciones y deben cumplirse las leyes de conservación, como la conservación de la carga eléctrica, el momento relativista, la energía, etc. En un vértice de un diagrama de Feynman, las partículas emergentes pueden ser diferentes a las entrantes (algo que es el pan nuestro de cada día en los aceleradores), pero en QED solo aparecen vértices de tres partículas: dos con carga eléctrica y un fotón. En nuestro ejemplo vemos dos. Las líneas onduladas representan partículas virtuales portadoras de la fuerza que, en la interacción electromagnética, son fotones y no son directamente observadas.

En general, no solo en QED, podemos entender cualquier interacción entre partículas como el paso de un estado inicial a un estado final. Conociendo qué tipo de interacciones se dan entre las partículas y las leyes de conservación que satisfacen, la

teoría cuántica de campos nos permite calcular las probabilidades de que, a partir de la interacción de unas partículas iniciales, acabemos en otras partículas finales, que en general no serán las mismas. Volvamos a nuestro ejemplo de la repulsión de dos electrones. Para calcular la probabilidad de que las cosas sucedan como describe el diagrama de Feynman necesitamos hacer una serie de operaciones con la siguiente expresión matemática: $(O_1 ie\gamma^\mu I_1)\left(\frac{-ig_{\mu\nu}}{p^2}\right)(O_2 ie\gamma^\nu I_2)$, donde la contribución de un electrón antes de la colisión corresponden al símbolo I y la de después de la colisión a la letra O. El subíndice indica si se trata del electrón 1 o 2. La contribución del fotón se corresponde con el cociente central de la ecuación. La genialidad de los diagramas de Feynman es que son una manera precisa y sencilla de representar este tipo de expresiones, como podemos deducir de la siguiente tabla:

Objeto	Diagrama	Símbolo
Electrón incidente		I_1 o I_2
Electrón resultante		O_1 o O_2
Fotón		$\frac{-ig_{\mu\nu}}{p^2}$
Interacción /Vértice		$ie\gamma^\mu$ o $ie\gamma^\nu$

Cada elemento del diagrama se traduce en parte de la expresión matemática. ¿Y ya está, es así de fácil? Lamentablemente solo hemos descrito una posible forma de pasar del estado inicial de nuestros electrones al estado final. Pero en realidad, la teoría cuántica de campos permite que esto suceda de una infinidad de maneras. En nuestro ejemplo, los siguientes diagramas de Feynman también son compatibles con una repulsión entre dos electrones:



En el primero, uno de los electrones resultantes emite un fotón virtual y lo reabsorbe. En el segundo diagrama, uno de los electrones emite un fotón antes de la colisión y lo reabsorbe posteriormente. Y por último, en el tercer diagrama, se crean y destruyen partículas virtuales intermedias. Así, que simplemente añadiendo un nuevo vértice de interacción, algo que es posible según la cuántica de campos, creamos un nuevo diagrama y el resultado final es el mismo. Hay infinitas posibilidades que nos llevan al resultado final de la repulsión de dos electrones. Pero, si hay tantos diagramas posibles, ¿cuál es el que realmente ocurre? Pues según la cuántica, simplemente no podemos saberlo, porque en cierto modo, todos los diagramas ocurren a la vez. La teoría nos dice que para determinar la probabilidad de que se produzca el estado final a partir del inicial debemos sumar las contribuciones a la probabilidad de todas las formas posibles de hacerlo, de todos los posibles diagramas de Feynman.

Cada uno de los posibles procesos que nos lleva al mismo resultado final está asociado a una integral de camino, que utiliza expresiones matemáticas como la mostrada anteriormente. Para computar la probabilidad exacta de alcanzar este estado final debemos sumar infinitas contribuciones, infinitos diagramas de Feynman. Afortunadamente, la emisión de un fotón en este proceso tiene una probabilidad de ocurrir de un 1%, mientras que la emisión de dos fotones tiene una probabilidad de un 0,01%. Y esta probabilidad va disminuyendo más y más mientras más fotones virtuales intervengan; es decir, mientras más vértices tenga nuestro diagrama. De modo que nuestro primer diagrama, el más simple, es el más importante.

La teoría de perturbaciones nos enseña cómo es posible resolver ecuaciones muy complicadas, en nuestro caso el cálculo de la probabilidad de un suceso, simplificándolas para obtener una solución tan aproximada como queramos. Es algo parecido a lo que hacemos al tomar como aproximación a una expresión matemática un truncamiento de una serie de Taylor. Nos quedamos con los primeros términos más relevantes. En nuestro ejemplo, el intercambio de un único fotón es la contribución dominante y, solo en el caso de que necesitemos cálculos más precisos, necesitaremos tener en cuenta otros diagramas. La ventaja de los diagramas de Feynman sobre el resto de formulaciones teóricas es que es relativamente fácil dibujar todos los posibles diagramas con dos o tres vértices, por ejemplo. Aun así, los físicos de partículas tienen que evaluar unos cientos de diagramas de Feynman para obtener aproximaciones a bajo orden y decenas de miles para ir algo más allá en la precisión. Algo que sin la inestimable ayuda de los ordenadores simplemente sería imposible hoy día.

La teoría cuántica de campos en su aplicación a la física de las partículas elementales es la teoría mejor confirmada experimentalmente de la historia. Y en QED la concordancia entre resultados teóricos y medidas experimentales sobrepasa con mucho lo espectacular. Pero los físicos no paran y trabajos recientes apuntan a que, tal vez, la época de los diagramas de Feynman toque pronto a su fin. Herramientas de la teoría de los twistors, desarrollada por Roger Penrose en la década de los 70 y usada entre los teóricos de cuerdas recientemente, prometen una nueva simplificación espectacular de los cálculos. Quizás así los diagramas de Feynman desaparezcan de la práctica diaria de los físicos de partículas, pero ocuparán siempre un lugar destacado en la historia de la física.

Bibliografía:
David Kaiser, "La física y los diagramas de Feynman". Investigación y ciencia, septiembre de 2005.
Richard P. Feynman, "Electrodinámica cuántica: la extraña teoría de la luz y la materia". Alianza Universidad, 1992.
Jesús Navarro Faus, "Feynman: Los caminos cuánticos". Nivola Libros y Ediciones, 2007.

Bartolo Luque
Doctor en Física
E.T.S.I. Aeronáuticos, Universidad Politécnica de Madrid

Capítulo 42
¿Qué es el principio de incertidumbre de Heisenberg?
(Por Ángel S. Sanz)

42.1 Introducción.

Cuanto más precisa sea la medida de la posición de una “partícula” {1}, más imprecisa será la medida de su momento lineal (el producto de su velocidad por su masa), y viceversa. Posiblemente ésta sea la versión más popular (ver Fig. 1) del *principio de incertidumbre* o *indeterminación* {2} de Heisenberg [1], quien descubrió que ésta era una de las consecuencias directas de la formulación o representación matricial de la mecánica cuántica, desarrollada por él mismo, Kramers, Born y Jordan en 1925 [2]. Formalmente, se expresa mediante la desigualdad

$$\Delta x \cdot \Delta p_x \geq \frac{h}{4\pi} \qquad (1)$$

donde Δx y Δp_x cuantifican, respectivamente, las incertidumbres asociadas a la medida de la posición y el momento de la partícula, y $h = 6.62607 \times 10^{-34}$ J·s es una constante física universal definida por Planck como el *quantum de acción* (*Wirkungsquantum*). Físicamente, la presencia de esta constante tiene como consecuencia la imposición de una limitación intrínseca a nuestro conocimiento y concepción del mundo cuántico, algo bastante alejado de la percepción que tenemos de los objetos cotidianos.



Fig. 1. *Sello conmemorativo alemán en el que aparecen Heisenberg y la relación de incertidumbre que lleva su nombre (Heisenbergsche Unschärferelation).*

El concepto de *quantum* fue introducido por Planck cuando buscaba una explicación para el espectro de la radiación que emite o absorbe un cuerpo negro [3], un problema que habían intentado resolver sin éxito eminentes físicos ya en el siglo XIX, como Rayleigh, Jeans o Wien a partir de la física conocida en ese momento, en particular, la teoría del electromagnetismo de Maxwell. A diferencia de ellos, Planck planteó la posibilidad de que la transferencia de energía desde un sistema a otro no se transmitiese de forma continua (es decir, no se puede transferir cualquier cantidad de energía), sino discreta, en forma de “paquetes” idénticos de un determinado tamaño a los que denominó *quanta* de energía.

$$E = h\nu = \frac{hc}{\lambda} \qquad (2)$$

La energía de cada una de estas unidades quedaba fijada por la frecuencia de la radiación asociada, donde ν y λ denotan la frecuencia y longitud de onda de la radiación (cuanto más rápida sea la oscilación de la radiación, menor será su longitud de onda, y viceversa). Esta idea tan innovadora dio lugar a lo que hoy se conoce como vieja teoría del *quantum*, empleada exitosamente por Einstein para suministrar una explicación del efecto fotoeléctrico mediante el concepto de *quantum de luz* (*Lichtquant*) y, una década más tarde, por Bohr como uno de los elementos centrales de su modelo atómico, para explicar por qué los átomos emiten o absorben luz solo con determinadas frecuencias o energías. Posteriormente, la constante de Planck se convertiría también en uno de los elementos clave de la mecánica cuántica, cuya aparición en 1925 supuso la superación, substitución, de la teoría del *quantum* de Planck.

Al igual que el valor de la velocidad de la luz en el vacío es un factor limitante de cuán veloz puede ser un objeto, la constante de Planck establece un límite al valor de la acción asociada a un sistema, lo que a su vez se traduce en la aparición de relaciones de incertidumbre para cualesquiera dos magnitudes cuyo producto tenga dimensiones de acción (energía por tiempo). Éste es el caso, por ejemplo, de la desigualdad (1), donde el producto de las dispersiones en los valores de la posición y el momento tiene precisamente esas dimensiones. Como consecuencia, mientras que la medida de la posición y la velocidad del péndulo de un reloj de pared no plantea ningún inconveniente (la acción asociada es muchísimo mayor que el valor de h y, por tanto, a efectos prácticos podemos asumir que $\Delta x \cdot \Delta p_x \approx 0$), el caso cuántico análogo requerirá de la realización de dos medidas (experimentos) independientes debido a la imposibilidad de poder determinar ambas cantidades con uno solo.

Las implicaciones de las relaciones de incertidumbre constituyen posiblemente el mayor inconveniente conceptual con el que nos encontramos cuando abordamos la mecánica cuántica y, por tanto, cuando intentamos comprender el comportamiento de los sistemas cuánticos, pues es imposible determinar con total precisión el valor de dos variables complementarias (formalmente, variables canónicamente conjugadas) cualesquiera que sirvan para especificar inequívocamente el estado de la partícula. La posición y el momento son un ejemplo; pero el mismo problema se nos plantea cuando intentamos especificar al mismo tiempo las tres componentes del momento angular o del espín de un sistema (o su estado de polarización, si se trata de un fotón), la energía asociada a su estado cuántico y su tiempo de vida media o, en el caso de sistemas de muchas partículas idénticas (fotones o electrones), el número de éstas y la fase que caracteriza a su estado, por ejemplo.

En este capítulo abordaremos una serie de aspectos asociados a las relaciones de Heisenberg, que nos permitirán comprender tanto su conexión con la teoría formal, como sus implicaciones físicas. Así, en la sección 42.2 se ofrece, sin entrar en detalles matemáticos, una visión de su procedencia dentro del marco de la propia teoría, tanto a partir de la formulación matricial de Heisenberg, como de la ondulatoria de Schrödinger. Para ilustrar el significado de la relación (1), en la sección 42.3 se describe un experimento mental (*Gedankenexperiment*) ideado por Heisenberg con tal propósito, el denominado *microscopio de Heisenberg* [4], y se introduce la noción de *complementariedad* o *principio de complementariedad* de Bohr. Para concluir, en la sección 42.4 se discutirán las implicaciones fundamentales a nivel interpretativo, principalmente dentro del contexto del concepto de complementariedad, la teoría de la medida [5], la denominada paradoja EPR y el entrelazamiento cuántico [6,7], así como experimentos actuales relacionados.

42.2 Formalismo mecano-cuántico y relaciones de incertidumbre.

Las relaciones de incertidumbre son un resultado directo del formalismo matemático de la mecánica cuántica, razón por la cual se podría considerar "dudoso" elevarlas a la categoría de principio fundamental, ya que estos, generalmente, son fundamento y no consecuencia. Ofrecer un análisis detallado de cómo se obtienen dichas relaciones excede el alcance de este capítulo, por lo que únicamente nos vamos a limitar a suministrar una serie de "pinceladas" que permitan visualizar cómo se deducen de la propia teoría y cuál es su significado físico. Para ello resulta instructivo emplear las dos formulaciones más estándar de la mecánica cuántica, la mecánica matricial de Heisenberg y la mecánica ondulatoria de Schrödinger, comenzando por la primera, más operacional y abstracta, y continuando por la segunda, más cercana a nuestra intuición.

El acercamiento de Heisenberg a la mecánica cuántica está basado en la utilización del álgebra matricial como pieza clave. Las matrices son objetos matemáticos que nos permiten organizar sistemáticamente, en forma de tabla, información referente a una cierta propiedad de un sistema físico. Esta información puede ser, por ejemplo, las diferentes energías (frecuencias) que originan y explican el espectro visible del átomo de hidrógeno. Esta idea tan aparentemente sencilla es, precisamente, lo que llevó a Heisenberg [2] al descubrimiento de la mecánica cuántica sin conocimiento previo de que en matemáticas ya existían tales objetos (en este sentido, hay que reconocer a Heisenberg el mérito de redescubrir todas las propiedades del álgebra de matrices, algo que, sin embargo, era conocido por los matemáticos de su época).

Para comprender la conexión entre la nueva mecánica descubierta por Heisenberg y de dónde surgen las relaciones de incertidumbre es preciso entender la forma de operar con matrices. Esto podemos verlo sin necesidad de conocimientos previos sobre álgebra matricial. En la física newtoniana, la evolución de un sistema se describe mediante trayectorias, es decir, asignando en cada instante un valor a la posición y el momento de la partícula. La variables posición, $x(t)$, y momento, $p(t) = m \cdot v(t)$, son ambas funciones dependientes del tiempo, t, que juega el papel de parámetro de evolución (o variable independiente). A un tiempo dado, podemos calcular el producto de ambas variables, para lo cual da exactamente igual multiplicar la posición por la velocidad, que al revés, es decir, $x(t) \cdot p(t) = p(t) \cdot x(t)$. Este producto es *conmutativo*, propiedad característica de la mecánica clásica.

A diferencia de las funciones, cuando tratamos con matrices, el producto ya no es conmutativo en general, es decir, $\hat{A} \cdot \hat{B} \neq \hat{B} \cdot \hat{A}$, a menos que ambas matrices describan propiedades del mismo tipo (por ejemplo, el momento de una partícula y su energía cinética). Esta anticonmutatividad se comprende fácilmente si cogemos, por ejemplo, un libro y lo rotamos 90º dos veces seguidas, tomando cada vez un eje de rotación diferente. Como se observa en la Fig. 2, dependiendo de qué eje de rotación se considere primero, se obtendrá un resultado u otro. En mecánica cuántica, las matrices se denominan operadores, porque actúan (operan) sobre el sistema y, al hacerlo, nos suministran información sobre la variable física a la que representan. Si no hay operador asociado, la magnitud no es un observable y, por tanto, no es un resultado experimentalmente accesible (no, al menos, de forma directa). Cuánticamente, posición y momento tienen asociados operadores y, por tanto, podemos medirlos, pero como representan propiedades diferentes, no conmutan:

$$\hat{X} \cdot \hat{P} - \hat{P} \cdot \hat{X} = i\hbar \qquad (3)$$

donde $\hbar \equiv h/2\pi$ es la constante reducida de Planck. Según esta relación, cuánticamente no existen estados en los que el sistema tenga una posición y un momento bien definidos al mismo tiempo, lo que implica que las incertidumbres de ambas variables serán no nulas, como establece la relación de incertidumbre desigualdad (1), que se obtiene, así, como consecuencia directa de la relación de no conmutación (3).

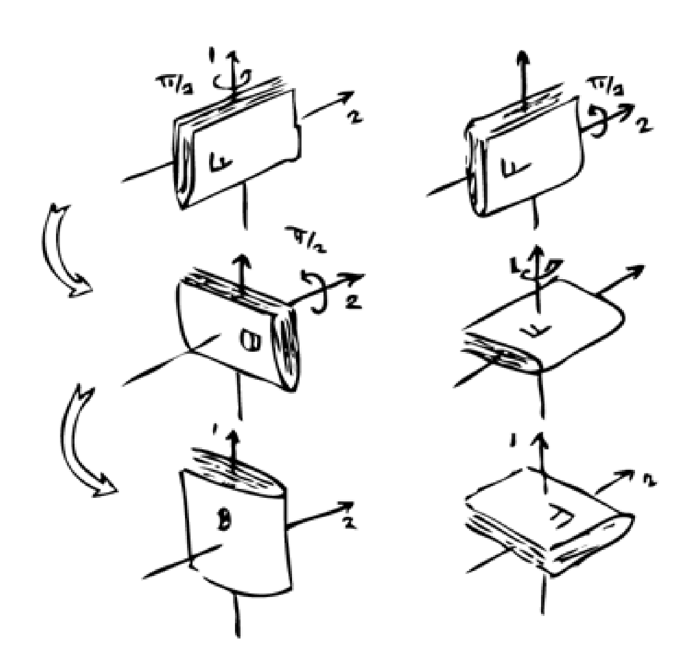


Fig. 2. *No conmutatividad de la rotación: la operación de rotar dos veces consecutivas un objeto a través de sus ejes de simetría no es conmutativa, pues el resultado final depende de qué eje de rotación se considera primero y cuál después. En este ejemplo, lo que hacemos es rotar dos veces un libro un ángulo de 90º tomando cada vez uno de sus ejes de simetría.*

Pese a describir también una física nueva, la formulación ondulatoria de Schrödinger no supone una ruptura total con la física clásica, o no, al menos, tan dramática, como la de Heisenberg. Ello se debe a que, de algún modo, parte de la física clásica, aunque no lo hace desde la formulación newtoniana, sino de la hamiltoniana. Ésta última constituye una generalización de la mecánica newtoniana a partir de la analogía entre acción mecánica y fase óptica, conocida como *analogía hamiltoniana*. Así, la mecánica ondulatoria describe la nueva física apoyándose en un lenguaje matemático (teoría ondulatoria) conocido para los físicos de la época, evitando el rechazo inicial que sí causó la mecánica matricial de Heisenberg. El punto de partida de Schrödinger es la hipótesis propuesta en 1925 por de Broglie, según la cual las partículas de materia pueden manifestar comportamientos de tipo ondulatorio (ver Capítulo 49), de igual manera que la luz puede hacerlo como una partícula (fotón), como propuso veinte años antes Einstein para explicar el efecto fotoeléctrico. Dentro de este contexto, la contribución fundamental realizada por Schrödinger fue encontrar la ecuación de movimiento que satisfacían las ondas o campos de materia, tomando como base la analogía hamiltoniana: hacia el primer tercio del siglo XIX, William Rowan Hamilton observó que en el límite de la óptica geométrica (longitudes de onda pequeñas), la descripción de las ondas admitía una descripción más simple en términos de rayos que satisfacían el principio del tiempo mínimo de Fermat. Lo que hoy se conoce como principio de mínima acción de Hamilton no es otra cosa que la versión mecánica del principio de Fermat; las trayectorias newtonianas, en virtud de la formulación hamiltoniana, son el equivalente mecánico de los rayos ópticos. Schrödinger generaliza esta idea al asumir que, cuando nos desviamos del límite de la óptica geométrica, las fases de las ondas asociadas a la luz son el análogo óptico directo de las fases de las ondas de materia de de Broglie. A partir de esta relación tan aparentemente directa y evidente se obtiene la ecuación de movimiento para dichas ondas, que hoy conocemos como ecuación de Schrödinger.

De acuerdo con la mecánica ondulatoria, el estado en el que se encuentra la partícula viene descrito por un campo denominado función de onda. Estos objetos

matemáticos son funciones que dependen de varias variables y parámetros. Ejemplos de función campo conocidos son la temperatura o la presión de los mapas meteorológicos: dadas unas coordenadas (expresadas en estos casos en términos de latitud y longitud por conveniencia) obtenemos el valor de temperatura o presión sobre el punto deseado; estos mapas (los valores de temperatura y presión) cambian con el tiempo, que es el parámetro de evolución. En el caso mecano-cuántico, la función de onda representa una amplitud de probabilidad, una función compleja, cuyo módulo elevado al cuadrado nos da el valor de la probabilidad con que podemos encontrar a una partícula en una determinada posición a un tiempo dado (o, en general, con cualquier otra propiedad física que defina el estado del sistema bajo estudio). Describir el estado de una partícula mediante una onda, con independencia de cuestiones metafísicas relacionadas con su interpretación, resulta muy interesante para el físico, ya que su análisis será exactamente equivalente al de cualquier onda, campo o señal (la terminología depende del área de la Física a que hagamos referencia, aunque el objeto en sí es siempre el mismo): en términos de análisis de Fourier. Una propiedad interesante de este tipo de análisis es que cualquier onda que dependa de una cierta variable, como, por ejemplo, la posición, tiene una representación dual en un espacio recíproco en términos de una variable conjugada asociada, como es el momento. Este momento, sin embargo, no tiene nada que ver con el momento canónico clásico, que es una variable independiente, sino que está intrínsecamente ligado al espacio recíproco de la variable de partida. Tal vez en el caso de posiciones y momentos esto no sea evidente, pero sí que lo es si pensamos que la transformada de Fourier de una función dependiente del tiempo es una función que depende de la frecuencia (o energía). Así, el momento al que se refiere la transformada de Fourier es en realidad una cantidad proporcional a una frecuencia espacial, es decir, $p = \hbar k$. De hecho, a partir de esa relación y de la que existe entre esta frecuencia y la longitud de onda, $k = 2\pi/\lambda$, se recobra inmediatamente la hipótesis de de Broglie

$$p = \frac{h}{\lambda} \qquad (4)$$

Pero no solo eso, sino que además los valores de las incertidumbres asociadas a esas posiciones y sus respectivos momentos (duales) satisfacen la relación (1), que puede ser obtenida de una forma relativamente sencilla a partir de teoremas del análisis de Fourier.

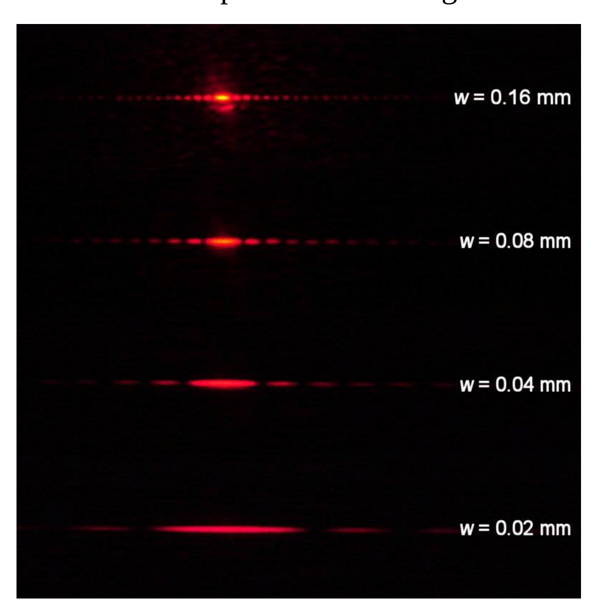


Fig. 3. Difracción de la luz producida por un haz láser rojo por rendijas de diversas anchura (decreciente de arriba abajo). A medida que decrece la anchura de la rendija, la difracción es mucho más importante (máximos más anchos).

Un experimento relativamente simple donde se observan las implicaciones de la relación de incertidumbre de forma inmediata es el de la difracción de un haz láser (luz procedente de un puntero láser rojo, por ejemplo) por una rendija practicada sobre un trozo de cartulina oscura. La difracción es un fenómeno distintivo de cualquier onda que consiste en que, cuando ésta atraviesa un obstáculo, muestra un patrón característico que nada tiene que ver con las zonas de sombra (de luz, sonido, telefonía, etc.) a que estamos acostumbrados en la vida cotidiana. En el caso del haz láser atravesando la rendija, dependiendo de la anchura de la rendija, se obtendrán

patrones como los que aparecen en la Fig. 3, con “manchas” de mayor amplitud a medida que la anchura de la rendija disminuye. Supongamos ahora que el patrón está formado por una gran cantidad de partículas cuánticas, ya sean fotones en el caso del láser o de electrones, átomos o moléculas, en otros casos. Consideremos una rendija de anchura w. La incertidumbre asociada a la onda que cruza la rendija es del orden de la anchura de ésta, $\Delta x \sim w$; es imposible saber por dónde pasará exactamente una partícula, solo sabemos que su posición estará en algún lugar dentro de ese rango. La porción de onda asociada (recordemos que adscribimos a la partícula una onda y no un punto localizado espacialmente) que atraviesa la rendija es una onda de perfil cuadrado, de valor distinto de 0 a lo largo de la rendija, por efecto de los bordes de ésta sobre la onda incidente. Físicamente, esto quiere decir que existe exactamente la misma probabilidad de que la partícula pase por cualquier punto de la abertura y que no hay posibilidad alguna de que lo haga por cualquier otro punto fuera de esa región. La transformada de Fourier es una función (onda) oscilante, cuyo argumento es $wp/\hbar$. El cuadrado de esta función es precisamente lo que se observa en la Fig. 3. La anchura del pico central de esta función (la distancia entre los mínimos que delimitan el máximo central de luz), que tomamos como incertidumbre del momento, es $\Delta p \sim h/w$. Por tanto, el producto de ambas incertidumbres reproduce la relación de Heisenberg directamente. Esto que hemos obtenido para la difracción de un haz láser se cumple escrupulosamente para cualquier tipo de onda y, en particular, para las ondas de la mecánica cuántica, describan éstas un electrón (partícula elemental sin estructura) o una gran biomolécula (partícula con una gran complejidad interna).

42.3 Complementariedad y el microscopio de Heisenberg.

La relación de incertidumbre de Heisenberg se haya íntimamente ligada a otro concepto o principio también central de la física cuántica: el *principio de complementariedad* de Bohr [8]. Según Bohr, la relación de incertidumbre no es sino una manifestación de la complementariedad cuántica, una idea o principio fundamental según la cual dos cualesquiera propiedades complementarias asociables a un sistema cuántico no pueden ser observadas al mismo tiempo en un experimento. Por ejemplo, un sistema cuántico no puede exhibir simultáneamente su carácter de partícula (o corpúsculo) y de onda (ver Capítulo 49 sobre el concepto dualidad onda-corpúsculo); en unos casos mostrará el primero y en otros el segundo, dependiendo ello de cómo esté diseñado el experimento que se vaya a realizar (ver discusión más detallada sobre los experimentos de la doble rendija de Young o de la elección retardada de Wheeler en el Capítulo 49). Igualmente, dado que la posición y el momento de un sistema cuántico son variables complementarias, no pueden ser determinadas en un mismo experimento con total precisión y, de ahí, que se observe una relación entre sus incertidumbres. Un ejemplo sencillo que ilustra cómo la complementariedad da pie al concepto de incertidumbre en la forma prescrita por Heisenberg, es decir, como una propiedad intrínseca del sistema bajo estudio, es el experimento mental que hoy conocemos como *microscopio de Heisenberg* [9], ideado por éste para explicar la física que entraña la desigualdad (1). La idea es muy simple. Supongamos que queremos observar la posición precisa de un electrón mediante un microscopio, como se ilustra en la Fig. 4. En principio, el electrón se encuentra a una distancia *d* del objetivo, sobre el eje óptico de esta lente, de diámetro D.

En esta configuración, electrón y objetivo forman un cono de abertura angular θ. Supongamos que los fotones con los que se ilumina al electrón para poder observarlo proceden de una fuente situada debajo del mismo, de manera que inicialmente no poseen componente transversal (su momento es perpendicular al plano del objetivo). El momento lineal de estos fotones es p_γ^0, que corresponde a una longitud de onda asociada λ_γ^f en virtud de la relación $p = h/\lambda$. Asumiendo que los fotones son altamente energéticos (rayos X o rayos gamma, por ejemplo), estos mostrarán su naturaleza corpuscular (ver Capítulo 49), como en el efecto fotoeléctrico. En tal caso, por efecto Compton, tras impactar sobre el electrón, los fotones serán dispersados en diversas direcciones, adquiriendo una cierta componente transversal (paralela al plano del objetivo) en su momento lineal. Igualmente, el electrón, que estaba inicialmente en reposo, gana cierta cantidad de momento, experimentando un desplazamiento a lo largo de la dirección del eje óptico (y también transversalmente). Dado que los fotones no disponían inicialmente de momento transversal, por conservación del momento lineal total del sistema electrón-fotón, el momento que ganan los fotones hacia un lado lo tienen que ganar también los electrones hacia el otro, de manera que el momento neto total quede cancelado. Así, el momento transversal de los electrones es

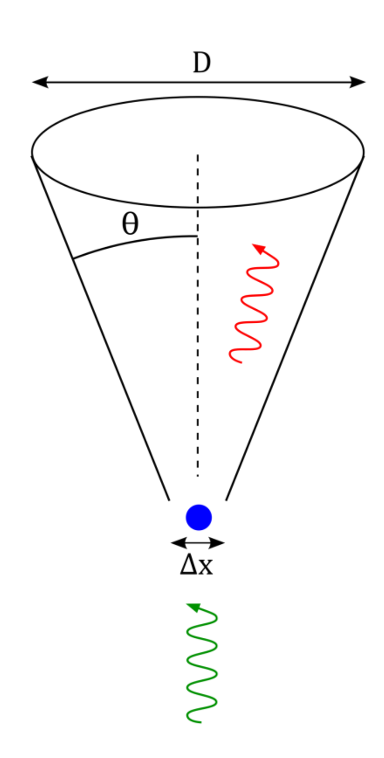


Fig. 4. Esquema general ilustrando la física del microscopio de Heisenberg. Solamente los fotones dispersados hacia el interior del cono determinado por el tamaño de la lente y su distancia focal nos pueden aportar información sobre la posición del electrón "observado".

$$p_{e,x} = p_e \sin\varphi = p_\gamma^f \sin\vartheta = \frac{h}{\lambda_\gamma^f} \sin\vartheta \qquad (5)$$

donde φ y ϑ son los ángulos de las trayectorias que siguen, respectivamente, el electrón y el fotón tras la colisión, es el ángulo de la trayectoria del fotón con respecto al eje óptico del sistema tras la colisión, p_e y p_γ^f son el momento final que adquieren, respectivamente, el electrón y el fotón, λ_γ^f es la longitud de onda del fotón tras la colisión, que viene dada por la expresión

$$\frac{1}{\lambda_\gamma^f} - \frac{1}{\lambda_\gamma^0} = \frac{h}{m_0 c}(1 - \cos\vartheta) \qquad (6)$$

siendo m la masa en reposo del electrón y c la velocidad de la luz en el vacío. El máximo ángulo de dispersión de los fotones que nos permite tener información sobre la presencia del electrón viene dado por la máxima abertura angular del cono imaginario antes mencionado. Cualquier fotón que salga dispersado con un ángulo mayor no llegará a nuestro ojo y, por tanto, no nos aportará información sobre el electrón. Ese ángulo implicará que el electrón, a su vez, ha ganado cierta cantidad de movimiento y, por tanto, su estado ya no es de reposo, como sucedía antes de la colisión. Dado que el

proceso de colisión es aleatorio, el electrón habrá ganado un momento cuyo valor estará acotado por el máximo momento hacia adelante y el máximo hacia atrás, es decir, estará comprendido en el rango

$$\Delta p_{e,x} = p_\gamma^f \sin\theta - p_\gamma^f \sin(-\theta) = \frac{2h\sin\theta}{\lambda_\gamma^f} \approx \frac{2h\theta}{\lambda_\gamma^f} \tag{7}$$

En la última expresión se ha considerado, por simplicidad, la aproximación paraxial o de ángulo pequeño, que permite aproximar el valor del seno (o la tangente) por su argumento ($\sin\theta \approx \theta$), ya que asumimos que la distancia de observación es bastante mayor que el diámetro del objetivo ($d \gg D$). El resultado expresado por la relación (7) nos indica que, en principio, cuanto mayor sea la longitud de onda de los fotones dispersados, más acotado estará el momento del electrón y, por tanto, menor será su incertidumbre. De este modo, si utilizásemos radiación de muy alta frecuencia (longitudes de onda grandes), conseguiríamos reducir en gran medida la imprecisión en el momento y tendríamos al electrón más localizado espacialmente, pues los posibles valores del momento con los que podría salir despedido éste a lo largo de la dirección transversal (en la que pretendemos tenerlo localizado) serían cada vez menores y más cercanos a cero.

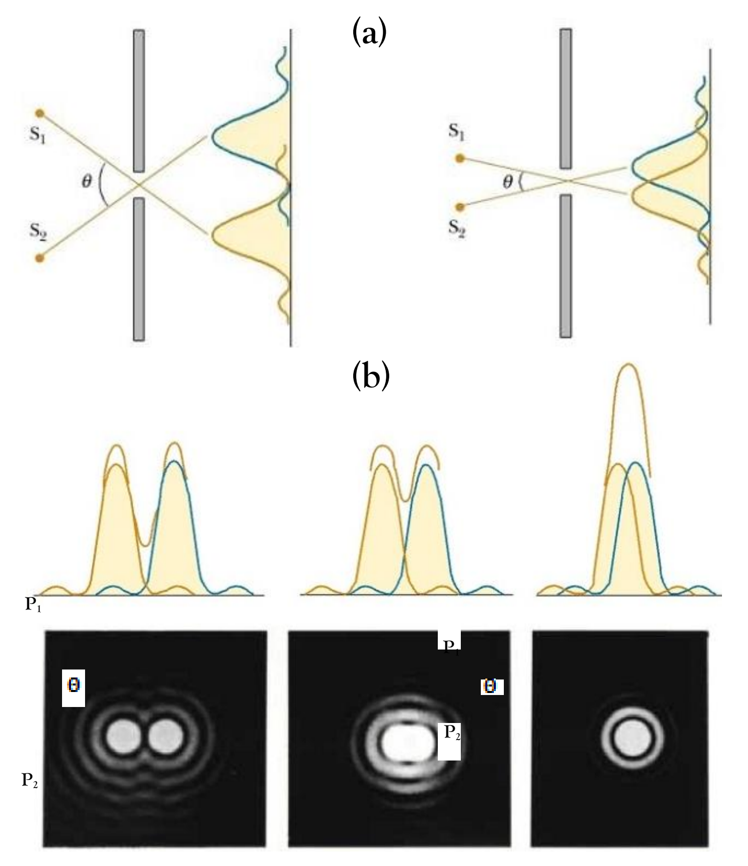


Fig. 5. (a) Sección transversal que muestra la difracción por una abertura circular de la luz emitida por dos fuentes puntuales o centros dispersores situados en posiciones alternativas P_1 y P_2, formando un ángulo θ. Cuanto menor es el valor de θ, más difícil se hace distinguir la procedencia del patrón de difracción. (b) De izquierda a derecha, patrones de difracción obtenidos con tres ángulos θ diferentes. El límite de la resolución vendría dado por la abertura angular que genera el patrón central, ya que aún podemos percibir la presencia de dos máximos (en el patrón de la derecha la superposición de los patrones nos impide saber si se trata de uno o de dos máximos).

Sin embargo, ¿significa eso realmente que podemos llegar a tener tan localizado al electrón como para determinar su posición precisa? Para contestar a esta pregunta, consideremos ahora el hecho de que, además de impactar sobre el electrón como si fuesen corpúsculos, los fotones también tienen asociada una naturaleza ondulatoria. En este caso, sabemos por la Óptica que, aparte de suministrarnos una imagen del objeto, una lente también se comporta como si fuese una abertura, es decir, difracta la luz que pasa a través de ella. Esto es lo que da lugar al denominado límite por difracción en la resolución de la lente (o, en general, del sistema óptico, como sería el caso de un microscopio), como se muestra gráficamente en la Fig. 5 (o se vio en el caso de la rendija de la Fig. 3).

Para comprender mejor las implicaciones del límite por difracción y su conexión con la relación de incertidumbre, consideremos que la lente se comporta, efectivamente, como una abertura circular, dando lugar a una cierta distancia el característico patrón de difracción, formado por círculo de luz, denominado disco de Airy, rodeado por una serie de anillos concéntricos de menor intensidad. La fuente de luz, en nuestro caso, es un centro de dispersión puntual, el electrón: cuando el fotón llega, la onda asociada empieza a formar frentes de onda del mismo modo que sucede cuando arrojamos una piedra a un estanque de agua. Supongamos que P_1 y P_2 son dos posiciones opuestas con respecto al eje óptico en las que el electrón podría estar localizado (asumimos, por hipótesis, que no estuviese inicialmente localizado). Como se muestra en la Fig. 5(a), se producirán dos patrones de difracción simétricos con respecto al eje óptico, uno asociado a cada posición. Si solo tuviésemos un centro dispersor situado sobre el eje óptico, el primer mínimo de patrón de difracción correspondiente, es decir, la primera zona de obscuridad que separa el disco de Airy del primer anillo concéntrico, ocurre cuando se cumple la condición

$$\frac{\pi D}{\lambda_\gamma^f} \sin \theta_d \approx 1.22\pi \tag{8}$$

donde θ_d es el ángulo subtendido desde el eje óptico a la posición de esa primera zona de obscuridad. En condiciones de paraxialidad, este ángulo es

$$\theta_d \approx 1.22 \frac{\lambda_\gamma^f}{D} \tag{9}$$

que además nos indica el tamaño (angular) del radio del disco de Airy. Según el criterio de Rayleigh, comúnmente utilizado para determinar el poder resolutivo de un sistema óptico afectado por difracción, para poder discernir si se trata de dos objetos diferentes o de uno solo, el ángulo Θ definido por las posiciones P_1 y P_2 con respecto al centro de la abertura debe de ser tal que los centros de los correspondientes patrones estaran, como mínimo a una distancia igual al radio del disco de Airy, es decir, $\Theta \approx \theta_d$ (ver Fig. 5(b)).

Conforme a lo anterior, para poder resolver las imágenes que generan los fotones que alcanzan al electrón, la posición de éste debe estar comprendida, como poco dentro del rango

$$\Delta x \approx \Theta d \approx 1.22 \frac{\lambda_\gamma^f d}{D} \tag{10}$$

Ahora, según se ha comentado antes, solo los fotones cuyo ángulo de salida sea igual o inferior a $\theta = D/2d$ podrán pasar a través de la lente. Por tanto,

$$\Delta x \approx 1.22 \frac{\lambda_\gamma^f}{2\theta} \tag{11}$$

de manera que lo que se obtiene ahora es que, cuanto menor sea la longitud de onda (radiación más energética), menos imprecisión tendremos y, por tanto, más localizado espacialmente estará el electrón. Es decir, llegamos justo a la conclusión contraria que

en el caso anterior. Nos encontramos con que no se puede variar la longitud de onda de la luz con que iluminamos al electrón sin redundar en perjuicio de una de las dos medidas. Ambas se hayan íntimamente relacionadas: si queremos localizar al electrón restringiendo su movimiento, incrementaremos la imprecisión asociada a la distorsión que causa la difracción; si queremos reducir el efecto de la difracción, es el movimiento del electrón lo que imposibilita su localización. Por ello, para encontrar un cierto consenso, lo que hacemos es calcular su producto, que resulta ser

$$\Delta x \cdot \Delta p_x \sim h \qquad (12)$$

que, aunque es independiente de la longitud de onda, nos deja ligadas ambas cantidades a través de la constante h. Este resultado nos indica que no podemos hacer ambas imprecisiones tan pequeñas como queramos y, por tanto, medir "la posición del electrón". Como señala Heisenberg tras introducir este experimento mental [4], siempre es posible alegar objeciones al mismo con el propósito de eludir dificultades en el proceso de medida y, por tanto, evitar la relación de incertidumbre. Sin embargo, se demuestra que éste es ineludible, ya que todas y cada una de esas objeciones añaden, a su vez, nuevas cuestiones que acaban redundando de una forma u otra en la relación de incertidumbre. Por ejemplo, se puede querer evitar el inconveniente de la incertidumbre en el camino del fotón desviado tras la colisión utilizando un microscopio móvil, lo cual conlleva al problema de la posición del propio microscopio y, de ahí, volveríamos de nuevo a la ecuación (12).

42.4 Interpretación de la mecánica cuántica y paradoja EPR.

Cuando situamos la relación de incertidumbre de Heisenberg dentro del contexto de la comprensión que la Física nos brinda del mundo que nos rodea, es decir, de cómo nos permite comprender y explicar los fenómenos que encontramos en la Naturaleza, ciertamente plantea un importante reto conceptual. A diferencia de los modelos propuestos con anterioridad a la mecánica cuántica, en los que existe una distinción o separación clara entre lo observado y el observador, ahora nos encontramos con que tal separación no existe en absoluto. Como hemos visto, si intentamos medir el sistema, lo estaremos influenciando de tal manera que no podremos conocer su estado antes de realizar el proceso de medida, pues éste entraña ya una perturbación del mismo. Aunque esto lo hemos discutido en el caso particular de la posición y el momento, la situación se hace extensiva a cualquier par de variables conjugadas que sirvan para especificar el estado del sistema (en el Capítulo 49 volveremos a esta cuestión precisamente dentro de la dicotomía onda-corpúsculo). Esta conclusión tan perturbadora para los físicos de principios del siglo XX es lo que llevo a lo que se conoce como interpretación ortodoxa o de Copenhague de la mecánica cuántica tras la 5ª Conferencia Solvay [10,11], en 1927, cuyo referente principal es Bohr y que, a día de hoy, sigue siendo la que aún se mantiene y utiliza para enseñar mecánica cuántica en las aulas. En su visión más radical, lo que establece es que hasta que no hay medida no hay sistema, es decir, que aunque sepamos que hay un objeto ahí, tan desconocido es para nosotros su estado, que a efectos prácticos es como si dicho sistema no existiese.

Obviamente, esa ruptura radical con la visión más materialista de la Física anterior, una ruptura que tiene cierta conexión con el neopositivismo del Círculo de Viena (la Segunda Conferencia Internacional por la Unidad de la Ciencia tuvo lugar precisamente en Copenhague, sobre el "Problema de la Causalidad", en 1936,

auspiciada por Bohr [12]) y el resurgimiento del idealismo alemán, principalmente durante los años 20 del siglo XX, no dejó indiferente a otro grupo de físicos de la época. Los físicos seguidores de esta otra corriente, denominada realista y liderada por Einstein, se mostraban escépticos y perplejos ante la posibilidad de que, en última instancia, la Física no nos permitiese acceder al mundo físico sin perturbarlo y, por tanto, consideraban que la teoría cuántica no podía ser una teoría completa, ya que no permite especificar completamente todas las propiedades que describen un objeto, lo que Einstein denominaba la "existencia real" del objeto. Esta perplejidad y, al mismo tiempo, preocupación queda bastante bien reflejada cuando en el transcurso del regreso a casa Einstein pregunta a Pais si realmente creía que la Luna solo existía cuando la miraba [13]. Igualmente, Schrödinger ponía de relieve esta ambigüedad a través de su famosa paradoja del gato en la caja, según la cual el estado del gato se haya indefinido, entre vivo y muerto, hasta que se abra la caja y se realice la observación (medida) por su acoplamiento al detonante del proceso que puede acabar su vida, un vial con un material radiactivo, cuya probabilidad de desintegrarse y activar el mecanismo viene descrita por la mecánica cuántica.

Como respuesta, Einstein, Podolsky y Rosen publicaron un trabajo en 1935 [14] en el que proponían como solución al dilema planteado por el principio de incertidumbre considerar dos sistemas cuánticos altamente correlacionados (entrelazados, como denominó Schrödinger al fenómeno [15], también en 1935), de manera que sobre una de las partículas se pudiese determinar su posición y, a partir de la otra, su momento, y viceversa, violando así la relación de incertidumbre. Acto seguido, Bohr demostró [16] que el razonamiento era erróneo y que, por tanto, la relación de incertidumbre no se violaba. Sin embargo, hoy día el experimento mental propuesto por Einstein, Podolsky y Rosen, denominado paradoja EPR, se ha convertido en una de las ideas más influyentes de la Física, siendo el germen de lo que ya se conoce como Segunda Revolución Cuántica y las incipientes tecnologías cuánticas [17, 18, 19]. El motivo es que introduce el concepto de no localidad, es decir, la interacción a larga distancia entre dos sistemas cuánticos como concepto clave (recuérdese que en la física anterior, la intensidad de la interacción entre dos sistemas disminuye con la distancia).

Hoy día las relaciones de incertidumbre siguen siendo inspiración para multitud de trabajos y experimentos por sus implicaciones sobre nuestra percepción del mundo cuántico. Se intentan diseñar experimentos que contravengan estas relaciones o, al menos, nos muestren vías que, como el experimento EPR, nos indiquen qué es lo que habría que hacer. Sin embargo, los resultados han sido negativos. Lo más cercano, tal vez, haya sido el experimento que Steinberg llevó a cabo hace unos años en la Universidad de Toronto [20], en el que se midieron al mismo tiempo (dentro de un mismo experimento) posiciones y momentos sin violar la relación de incertidumbre de Heisenberg. Esto permitió observar (o inferir) los dos aspectos complementarios de un sistema cuántico: onda y corpúsculo. Ésta es una historia que continúa en el Capítulo 49, pero que nos introduce una pregunta interesante: ¿son realmente las relaciones de incertidumbre de Heisenberg un resultado intrínsecamente fundamental de la mecánica cuántica?

Notas:
{1} A lo largo del capítulo se va a utilizar el término "partícula" por simplicidad y también porque está bastante extendido en la literatura divulgativa referente al principio de Heisenberg. Sin embargo, conviene especificar que el término correcto es "sistema", que es más abstracto y hace referencia al grado de libertad (variable) o conjunto de ellos que sirven para caracterizar el objeto físico o fenómeno bajo

estudio. Cuando nos referimos al movimiento de una partícula de masa m sin estructura interna, uno y otro término se identifican y, para simplificar, hablamos de partícula. Sin embargo, estamos describiendo la vibración o rotación del objeto, es decir, de una partícula con una estructura interna, el sistema viene descrito por unas variables que no permiten identificar a éste con dicha partícula, aunque tanto vibración como rotación tendrán asociadas unas relaciones de incertidumbre.
{2} En castellano se utilizan indistintamente los términos "incertidumbre" e "indeterminación" para hacer referencia al principio de Heisenberg, al igual que en alemán, donde se emplean las palabras "*Unschärfe*" y "*Umbestimmtheit*". Para simplificar, aquí emplearemos el término "incertidumbre", de uso bastante extendido debido a la influencia del inglés, donde se utiliza la palabra "*uncertainty*". Sin embargo, en vez de la palabra "principio" se utilizará "relación", igual que en alemán ("*Unschärferelation*" o "*Unbestimmetheitsrelation*"), ya que el concepto "principio" transciende el significado que se le da al resultado obtenido por Heisenbeg, que, como se indica en este capítulo, es una consecuencia directa de la estructura matemática de la mecánica cuántica y, por tanto, no puede considerarse como un principio físico fundamental, como lo son la complementariedad, la naturaleza estadística o probabilista intrínseca de la mecánica cuántica o el principio de exclusión de Pauli, que no son consecuencia, sino fundamento de la propia teoría, que se construye y desarrolla en torno a ellos.

Bibliografía:

[1] "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik", W. Heisenberg, *Z. Phys.* **43**, 172-198 (1927).
[2] "Sources of Quantum Mechanics", B. L. van der Waerden (Dover, New York, 1967).
[3] "The Theory of Heat Radiation", M. Planck (Dover, New York, 1959).
[4] "The dreams that stuff is made of", S. Hawking (Running Press, Philadephia, 2011).
[5] "Quantum Theory and Measurement", J. A. Wheeler and W. H. Zurek (Princeton University Press, Princeton, NJ, 1983).
[6] "The Age of Entanglement", L. Gilder (Alfred A. Knopf, New York, 2008).
[7] "La realidad cuántica", A. Cassinello y J. L. Sánchez Gómez (Crítica, Barcelona, 2012).
[8] "The quantum postulate and the recent development of atomic theory", N. Bohr, *Nature* **121**, 580-590 (1928).
[9] "The Physical Principles of the Quantum Theory", W. Heisenberg (University of Chicago Press, Chicago, 1930), chaps. 2 y 3 (re-impreso en la Ref. 4); "Quantum Theory", D. Bohm (Dover, New York, 1951), chap. 5.
[10] "The Solvay Councils and the Birth of Modern Physics", P. Marage and G. Wallenborn (Springer, Basilea, 1999).
[11] "Quantum Mechanics – Historical Contingency and the Copenhagen Hegemony", J. T. Cushing (University of Chicago Press, Chicago, 1994).
[12] "The Vienna Circle – Studies in the Origins, Development, and Influence of Logical Empiricism", F. Stadler (Springer, Heidelberg, 2015).
[13] "Subtle is the Lord – The Science and the Life of Albert Einstein", A. Pais (Oxford University Press, Oxford, 2005).
[14] "Can quantum-mechanical description of physical reality be considered complete?", A. Einstein, B. Podoslky and N. Rosen, *Phys. Rev.* **47**, 777-780 (1935).
[15] "Can quantum-mechanical description of physical reality be considered complete?", N. Bohr, *Phys. Rev.* **48**, 696-702 (1935).
[16] "Discussion of probability relations between separated systems", E. Schrödinger, *Math. Proc. Cambridge Phil. Soc.* **31**, 555-563 (1935); "Probability relations between separated systems", E. Schrödinger, *Math. Proc. Cambridge Phil. Soc.* **32**, 446-452 (1936).
[17] "The New Quantum Age – From Bell's Theorem to Quantum Computation and Teleportation", A. Whitaker (Oxford University Press, Oxford, 2012).
[18] "Tecnologías cuánticas", J. J. García Ripoll, *Acta Científica y Tecnológica* **25**, 2-7 (2015).
[19] "Hacia un mundo cuántico", A. S. Sanz, *Acta Científica y Tecnológica* **26**, 18-24 (2016).
[20] "Observing the average trajectories of single photons in a two-slit interferometer", S. Kocsis, B. Braverman, S. Ravets, M. J. Stevens, R. P. Mirin, L. K. Shalm and A. M. Steinberg, *Science* **332**, 1170-1173 (2011).

Ángel S. Sanz
Doctor en CC Físicas
Profesor Departamento de Óptica, Facultad de CC Físicas,
Universidad Complutense de Madrid

Capítulo 43
¿Qué es la paridad?
(Por Belén Valenzuela Requena)

Asimov nos contó en 1973 que los objetos que poseían la simetría de paridad (P) eran idénticos a su imagen en el espejo. La letra “x” tiene simetría especular mientras que la “p” no tiene simetría especular: la imagen especular de “p” es “q”. También nos contó cómo definir la paridad en los procesos de partículas subatómicas y así decidir si un proceso cumple esta simetría o no.

Pero ¿qué se entiende por simetría? La simetría siempre ha fascinado al ser humano, la simetría de las flores, de los copos de nieve, del panal de las abejas, nosotros mismos somos casi simétricos. Podemos definir simetría como una transformación en la que haces a algo y que cuando terminas de hacerla está igual que antes. Así, por ejemplo, si a un objeto cuadrado lo rotamos 90º no lo podemos distinguir del objeto original. El “algo” en nuestro caso sería las leyes de la física. Así, las leyes de la física tienen la simetría de la paridad, si cuando realizamos la transformación de la paridad cambiando sus coordenadas por sus opuestas, las leyes permanecen invariables. Asimov también nos habló de otras dos transformaciones que se pueden aplicar a las leyes de la física: la conjugación de carga C en la que las partículas se transforman en sus contrarias, antipartículas, y la inversión temporal T en la que se invierte el sentido del tiempo como si viéramos una película marcha atrás.

¿Por qué los físicos se preocupan tanto de las simetrías de las leyes? Hasta Einstein las leyes de la física tenían que cumplir las simetrías que observamos en el Universo y eso ya es suficientemente importante. Pero Einstein hizo una revolución en el pensamiento [1]. Él construyó su teoría de la relatividad *requiriendo que cumpliera ciertas simetrías*. Esta idea cuajó en la física moderna y la simetría es el mejor aliado de los físicos para construir las teorías que describen tanto la física subatómica, como un imán, como el Universo. En este sentido la simetría es bella y fructífera.

En particular, las simetrías C, P y T van a tener un papel fundamental en la comprensión de nuestro universo. En 1927 Eugene Wigner propuso que nuestro mundo cumple la simetría P con lo que la imagen en el espejo del mundo se comportaría de la misma manera. La simetría P estaba firmemente establecida para la interacción gravitatoria, la interacción electromagnética y la interacción fuerte pero a mediados de los años 50 Tsung Dao Lee y Chen Ning Yang pusieron de manifiesto que la situación era más confusa para la interacción débil. Estos investigadores contactaron con la investigadora Chien-Shiung Wu para proponerle experimentos que pudieran averiguar que ocurría con la interacción electrodébil. Wu, junto con la National Bureau of Standards demostraron en 1956 que la fuerza débil no cumplía la simetría P.

Para entender el experimento a nivel fundamental necesitamos adentrarnos brevemente en la interacción nuclear débil y en el Modelo Estándar (ME). El ME es la mejor teoría de la comprensión de nuestro Universo que tenemos hasta la fecha. En la época en que Asimov escribía su libro “100 preguntas básicas sobre la Ciencia” el ME se estaba construyendo. En este modelo existen dos tipos de partículas con distinta naturaleza, los fermiones y los bosones, estos últimos son las partículas intermediarias de las interacciones. Los fermiones más conocidos son los neutrones, protones, electrones y neutrinos [2] y el bosón más famoso es el fotón, que es la partícula intermediaria de la interacción electromagnética. Lo que nos interesa a nosotros para nuestra historia es que este modelo propone que los neutrones y protones no son partículas fundamentales si no compuestas y están formadas por partículas ahora si, fundamentales, llamadas quarks [3]. En particular un protón está formado por dos quarks up (u) de carga +2/3 y uno down (d) de carga -1/3, es decir, protón=uud con carga=+2/3+2/3-1/3=1, y un neutrón está formado por dos down y uno up, es decir, neutrón=ddu con carga=-1/3-1/3+2/3=0.

La interacción débil es la responsable de la desintegración radiactiva que da lugar a la fisión nuclear y con ello a la posibilidad de la bomba atómica. Es la única que puede cambiar los sabores de los quarks y transformar un quark down en uno up o viceversa. Así la interacción débil puede transformar un neutrón en un protón. En la reacción se emiten además electrones (e^-) y el antineutrino ($\underline{\nu}$, antipartícula del esquivo neutrino). De hecho, los neutrinos y antineutrinos interaccionan con la fuerza débil y no lo hacen ni con la fuerza fuerte ni con la electromagnética. Una posible reacción sería:

$d \text{ -> } u + \underline{\nu} + e^-$

Los quarks y los electrones están compuestos de una parte dextrógira y otra levógira. El experimento solo se podía entender si la interacción débil involucraba únicamente la parte levógira del quark y los antineutrinos eran dextrógiros sin correspondiente parte levógira. Los neutrinos dejan de manifiesto que la simetría de paridad está rota. En aquel momento en el que se creía que el neutrino no tenía masa el neutrino se describía como *un fermión de Weyl*. Este fermión cumplía las mismas ecuaciones que otros fermiones pero con masa cero lo que le daba unas propiedades especiales.

Lev Landau propuso entonces que la conjugación de la carga junto con la paridad –simetría CP- se tenía que conservar. Allí donde una partícula no era simétrica, su antipartícula tampoco lo era, pero de manera contraria. En palabras de Asimov, si una partícula era como “p”, su antipartícula era como “q”. En 1964 se observó indirectamente que esta simetría también estaba rota en algunos procesos que involucraban la interacción débil. No fue hasta 1990 que la rotura de esta simetría se pudo comprobar de forma directa. No voy a entrar en los detalles de los experimentos pero las consecuencias de este descubrimiento fueron enormes ya que entre otras cosas predijo la existencia de la tercera generación de quarks. Subsiguientes experimentos fueron encontrando los quarks puntualmente completando así el ME.

Con gran sorpresa en los años 90, Takaaki Kajita y Arthur McDonald, descubrieron que los neutrinos sí tienen masa. Por este descubrimiento han recibido el premio Nobel recientemente (año 2015). El ME fue pensado para describir neutrinos sin masa y actualmente se está trabajando intensamente para arreglar esta fisura. El descubrimiento dio lugar a innumerables experimentos sobre neutrinos. Entre las

consecuencias de este hallazgo, se piensa que la masa de los neutrinos tiene la llave (o una de ellas) para entender porque hay más materia que antimateria. Si hubiera habido tantas partículas como antipartículas en los primeros segundos después del Big Bang, se hubieran aniquilado entre ellas dejando solamente un eco de radiación. La simetría CP se debió de romper porque nosotros estamos por aquí.

La simetría que sí se conserva en el ME es la simetría CPT, es decir, si invertimos el tiempo, cambiamos las partículas por antipartículas y las coordenadas por sus opuestas, las leyes de la física permanecen iguales. Siguiendo la filosofía de Einstein, al imponer estas simetrías se deduce que el universo está compuesto de dos tipos de partículas, los fermiones y los bosones. Ha habido propuestas de violación de CPT en teorías que van más allá del ME pero la mayor parte de la comunidad científica piensa que la simetría CPT se tiene que cumplir también en modelos de gran unificación en los que además de la fuerza electromagnética, la débil y la fuerte se integra la interacción gravitatoria.

Curiosamente, muchos conceptos de la física de partículas son muy relevantes en la física de la materia condensada. La materia condensada es una física más cercana a nuestro día a día donde se pretende entender el comportamiento de las fases condensadas algunas bien conocidas como la fase líquida y la sólida y otras más exóticas como la fase magnética o la superconductora. En un principio parecería que esta física de baja energía (1eV) no tendría nada que ver con la física de partículas elementales que se realiza en el CERN (≥ TeV). Sin embargo, en los materiales se dan partículas efectivas que son bosónicas, o fermiónicas o incluso más exóticas. Las simetrías C, P y T tienen un papel fundamental en uno de los campos de investigación más activos de hoy en día, las fases topológicas de la materia. Estas fases además de ser una revolución en la comprensión de las fases de la materia podrían revolucionar nuestra manera de hacer computación. El año 2015 también fue un año especial porque experimentalmente se descubrieron los semimetales de Weyl [4]. Las partículas de Weyl no tuvieron futuro para explicar los neutrinos pero sí se han encontrado en materia condensada. Este descubrimiento pone de manifiesto una vez más la universalidad de los conceptos en física.

Puede que la próxima vez que te mires en el espejo te pase como Alicia y planees por los confines del Universo.

Notas:

[1] Standford Encyclopedia of Philosophy, Symmetry and Symmetry breaking. (2003, revisión 2013) http://plato.stanford.edu/entries/symmetry-breaking/

[2] Véase Capítulo 51

[3] Véase Capítulo 46

[4] Su-Yang Xu et al, “Discovery of a Weyl fermion semimetal and topological Fermi arcs”, Science 349.6248 (2015), pp. 613{617. issn: 0036-8075}.

Belén Valenzuela Requena
Doctora en Física
Científico Titular
Instituto de Ciencias de Materiales de Madrid ICMM CSIC

Capítulo 44
¿Qué tiene que ver nuestra sociedad con la superconductividad?
(Por Belén Valenzuela Requena)

Nuestra sociedad y la superconductividad tienen en común que son dos *estados emergentes.* Como veremos hay muchísimos ejemplos de estados emergentes.

Detrás del concepto de estados emergentes está la idea ya expuesta por Aristóteles de que el todo es más que la suma de las partes. Todos tenemos vivencias de esta reflexión. Sin embargo, esta idea se opone a como tradicionalmente nos hemos acercado a entender la Ciencia: estudiando los elementos que la constituyen. Claramente es difícil entenderlo todo a la vez y es muy racional ir estudiando parte por parte. Este método lo ideó Descartes, en su Discurso del Método. Descartes afirmaba que el mundo era una gran máquina y para entenderlo tendríamos que entender sus elementos constituyentes. En la búsqueda de los elementos constituyentes de la materia fuimos al átomo, de ahí a los electrones, neutrones y protones y por último a los quarks y al modelo estándar. [1]

Este pensamiento sigue influenciando nuestro día a día y así cuando tenemos una dolencia seria vamos al médico especialista. Si el diagnóstico no es claro nos pueden mandar al de digestivo y luego al endocrino y luego al neurólogo y luego vuelta a empezar y en la sala de espera nos preguntamos si no sería mucho más eficiente que hubiera una comunicación fluida entre ellos. Nuestro cuerpo está claramente conectado.

Y es que aunque se han resuelto muchos problemas con el método de Descartes, muchos fenómenos cercanos de nuestro día a día no pueden ser entendidos reduciéndolos a sus elementos constituyentes. Entendiendo un átomo de hierro no podemos entender el estado magnético, entendiendo una molécula de ADN no podemos entender la vida, entendiendo una neurona no podemos entender la consciencia y aún entendiendo a las personas no podríamos entender la sociedad. El magnetismo, la vida, la consciencia y la sociedad son estados emergentes que surgen debido a la interacción entre sus elementos constituyentes (el hierro, las moléculas, las neuronas y las personas) y la interacción con su entorno. El estado emergente es entonces una propiedad que *no existía* en sus elementos constituyentes y no se puede utilizar el método de Descartes para entenderlo.

¿Cómo abordar entonces este tipo de problemas? Mi campo de investigación es la física de la materia condensada que estudia las fases líquidas y sólidas de la materia. Uno de los mayores héroes de mi campo, Phil Anderson, premio Nobel de física, escribió en 1972 un artículo provocativo con el título “*More is different*“ donde explicaba el alcance y la potencia del concepto de emergencia. En este artículo Anderson argumentaba que la habilidad de reducir el todo a leyes fundamentales no implica la habilidad de empezar desde esas leyes y reconstruir el universo. Estas ideas

no eran nuevas en absoluto pero todavía la aproximación de Descartes era la dominante a la hora de abordar un problema en ciencia. Anderson argumentaba que en la física de la materia condensada hay conceptos y herramientas que nos permiten atacar estos problemas, en particular la teoría de las transiciones de fase.

La teoría de transiciones de fase no se limita a la materia condensada, también se da en otras ramas de la física. En general se da siempre que tengamos un sistema de partículas que interaccionan entre ellas y/o con su entorno. El reconocido físico Lev Landau contribuyó enormemente a construir esta teoría. Landau propuso que las transiciones pasan de una fase ordenada a una fase desordenada y a la transición la denominó zona crítica. El ejemplo típico de transiciones de fase es el paso de líquido (fase desordenada) a sólido (fase ordenada) de un material cuando disminuimos la temperatura. Si por ejemplo el material fuera la sal las partículas constituyentes serían el sodio y el cloro. Cuando se pasa al estado sólido las moléculas de la sal se ordenan de una forma cúbica. También disminuyendo la temperatura podemos pasar de un material no magnético (fase desordenada) a uno magnético (fase ordenada) a una cierta temperatura crítica. Aquí, las partículas constituyentes son los electrones que poseen un pequeño momento magnético denominado spin. En la fase ordenada del imán todos los momentos magnéticos se alinean en la misma dirección creando un campo magnético propio. Aunque a primera vista el magnetismo y las fases de la sal no tienen nada que ver, la zona crítica de las transiciones *se describe con las mismas ecuaciones*. A este hecho se le denomina "fenómeno de Universalidad", uno de los conceptos más bellos de la física. Lo más sobresaliente de la teoría es que aunque la interacción entre átomos de la sal o la interacción entre los electrones del material es de un alcance corto que involucra solo a los vecinos más próximos, en la zona crítica todas las partículas constituyentes están correlacionadas y sienten lo que hacen desde sus vecinas más próximas a las más distantes. Es lo que se denomina un fenómeno cooperativo. La zona crítica es invariante de escala con propiedades similares en todas las longitudes y puede dar lugar a los bellos fractales.

Mi ejemplo favorito de transición de fase en materia condensada es el paso de un material a un estado superconductor cuando bajamos la temperatura. Los superconductores tienen propiedades eléctricas y magnéticas muy exóticas. No presentan resistencia al paso de la corriente eléctrica con lo que no pierden energía y además expulsan el campo magnético. Esto último da lugar a la posibilidad de observar el divertido fenómeno de levitación de o bien un imán sobre un superconductor o viceversa. Como es de esperar las posibles aplicaciones de los superconductores son enormes: cables no disipativos, imanes superpotentes, trenes que levitan... [2] La transición de fase en la que se pasa de estado normal a estado superconductor se da a temperaturas críticas muy bajas o recientemente a presiones muy altas [3] lo que limita el alcance de las aplicaciones. Se está investigando muy intensamente para conseguir temperaturas críticas más altas y optimizar estos materiales. Además, todavía no se entiende porque algunos de los materiales son superconductores. Esto hace que la superconductividad sea un fenómeno aún más apasionante para los investigadores.

Igual que en el caso del magnetismo los protagonistas de la transición de fase al estado superconductor son los electrones. En un material los electrones que conducen la corriente eléctrica son compensados por los iones que se encuentran formando una red cristalina. A temperatura alta hay resistividad que se produce cuando estos electrones interaccionan entre si y con las vibraciones de la red iónica. A baja temperatura en la fase superconductora sin embargo, los electrones se emparejan formando los

denominados pares de Cooper y fluyen todos a una en una onda colectiva siguiendo las reglas de la física cuántica. De nuevo se pasa de una fase desordenada con resistencia a una fase ordenada que es una onda colectiva al bajar la temperatura.

Cuando lo que queremos describir son sistemas biológicos el problema es mucho más complejo. Una de las dificultades es que estamos tratando con sistemas de no-equilibrio, es decir, que evolucionan con el tiempo y tienen *historia*. Pero la historia no solo existe en los seres vivos, también en el mundo inanimado existen sistemas en no equilibrio. Como resultado de esa historia obtenemos los bonitos y diversos copos de nieve. Curiosamente algunas bacterias también se organizan en estructuras parecidas a copos de nieve. [4] ¡De nuevo con las mismas ecuaciones! Hay además multitud de ejemplos en los que se pueden utilizar conceptos de transición de fase en sistemas que evolucionan en el tiempo: en manadas de mamíferos, bancos de peces, enjambres de abejas, colonias de hormigas. En el siguiente artículo utilizan la teoría de Landau para entender el movimiento de las bandadas de estorninos. [5] Así, algunos aspectos de la física de equilibrio también son útiles en la física de no equilibrio.

En un rango completamente diferente hay propuestas en neurociencia en las que se argumenta que en nuestro cerebro los fenómenos cooperativos juegan un papel fundamental en determinar la dinámica del cerebro. Nuestro cerebro parece estar en la zona de critica en medio del orden y el desorden. De esta forma puede conseguir la dualidad esencial para que funcione: debe mantener algo de orden para asegurar un funcionamiento coherente y reproducible y a la vez permitir un cierto grado de desorden que permita flexibilidad. [6]

Como os estáis dando cuenta la lista de problemas que se puede tratar con estas técnicas es enorme y solo he escogido unos cuantos. ¿Y nuestra sociedad? ¿Tiene algo que decir la física en como construimos las ciudades, nos movemos en espacios abiertos, votamos y formamos alianzas y grupos? Efectivamente la física se puede usar para comprender ciertos aspectos del mercado económico y para revelar estructuras escondidas en redes sociales. En estos problemas se entiende como surge el conflicto y la cooperación.[4]

Fuera del ámbito del estudio de la socio-física la idea de emergencia ha calado como filosofía social. Por ejemplo, en el ámbito de la empresa se denomina sinergia. Stephen R. Covey argumenta en un influyente libro en EEUU [7] que las mejores empresas y familias son las que las relaciones humanas no se basan en:

1) ni tu, ni yo (ni como ni dejo comer)
2) tú o yo (típica competición)
3) yo te respeto y tu me respetas (suma de las partes)
En las mejores empresas se usa la cuarta solución:
4) más, más (el todo es más que la suma de las partes)

En el punto 4) la idea es que todos contribuyan para producir una lluvia de ideas de la que surja al final algo completamente nuevo y mejor.

Cuando escribí los agradecimientos de mi tesis utilicé la analogía de comparar la transición a un estado superconductor con la revolución en una sociedad. En vez de bajar la temperatura hasta la temperatura crítica se tendría que pasar a la presión social crítica en la que pasamos de un estado de personas individualistas a un estado colectivo

donde todos queremos conseguir lo mismo. Recientemente escuché una charla TED [8] al hilo de esta idea. David Steindl-Rast argumentaba que lo que queremos conseguir todos es ser felices y que la forma de hacerlo es ser agradecidos localmente, con nuestra familia, nuestros vecinos, nuestros compañeros, por las cosas que tenemos... Y según él habría un momento en el que se produjera una transición a un estado en el que todos seríamos felices, transición que será una verdadera revolución en la que no haya nadie por encima de nadie pero todos conectados en una red. Es una utopía sugerente y no pude evitar pensar en las transiciones de fase.

Notas:
[1] Véase Capítulo 46
[2] Véase por ejemplo: http://www.icmm.csic.es/superconductividad/
[3] A.O. Drozdov,M.I. Eremets, I.A. Troyan, V. Ksenofontov y S.I. Shylin, Nature 525, 73 (2015)
[4] Philip Ball, Critical Mass, how one thing leads to another. Editorial Arrow Books.
[5] Alessandro Attanasi et al. , Nature Physics DOI: 10.1038/NPHYS3035
[6] Dietmar Plenz, Critical Brain, *Physics* **6**, 47 (2013)
[7] Stephen R. Covey. Los siete hábitos de la gente altamente efectiva. Editorial Paidós.
[8] https://www.ted.com/talks/david_steindl_rast_want_to_be_happy_be_grateful?

Belén Valenzuela Requena
Doctora en Física
Científico Titular
Instituto de Ciencias de Materiales de Madrid ICMM CSIC

Capítulo 45
¿Por qué se habla de la semivida de un isótopo y no de su vida entera?
(Por Enrique Macías Virgós)

El término "semivida" (en inglés *half-life*, a veces traducido incorrectamente como "vida media") empezó a usarse para predecir el comportamiento de algunos isótopos radioactivos, pero puede aplicarse en muchos otros contextos.

La semivida es una noción matemática ligada a la idea de "probabilidad". Un átomo inestable puede liberar en cualquier momento algún tipo de partícula subatómica o radiación, convirtiéndose en otro elemento químico. El ritmo con que esto ocurre difiere de un elemento a otro. Además, es imposible predecir cuando va a tener lugar este "decaimiento radioactivo" para un átomo concreto; pero podemos usar técnicas estadísticas para estimar cómo va a comportarse una masa grande de átomos. En el modelo matemático que explica este proceso se observa que la cantidad de átomos radioactivos se reduce a la mitad siempre que transcurre un determinado tiempo. Este período de tiempo es la "semivida".

Por ejemplo, el carbono-14 se va convirtiendo en nitrógeno, con una semivida de 5730 años, y este dato puede servir para determinar la edad de un fósil. El ^{14}C se forma en las capas altas de la atmósfera por la acción de los rayos cósmicos; para simplificar, podemos suponer que su concentración se mantiene constante. Las plantas y los animales lo absorberán, en esa misma proporción, pero solo hasta su muerte, y la cantidad de isótopo que aún hay en el fósil nos dará una indicación precisa del tiempo que ha pasado. Por este método de datación recibió Willard F. Libby el Premio Nobel de Química en 1960 [1].

Otro ejemplo es el plutonio-238, que es de uso común para dar energía a algunas naves espaciales [2]; así, varias misiones Apolo dejaron en la Luna un "generador termoeléctrico de radioisótopo" (es decir, un aparato que convierte el calor producido por la radioactividad en electricidad) para permitir realizar experimentos durante años. Como el ^{238}Pu tiene una semivida de 88 años, si un generador de este tipo tiene una potencia de 100 Watios, al cabo de 40 años la potencia será algo más de 70W. El cálculo exacto requiere utilizar logaritmos, pero puede hacerse con una sencilla calculadora científica.

El modelo matemático que rige estos fenómenos aparece siempre que el ritmo al que cambia una cantidad sea proporcional a su valor actual. Algunas reacciones químicas, la altura del agua en una piscina que se está vaciando por el fondo o la evolución de algunos activos financieros siguen este modelo.

Un ejemplo típico en Farmacocinética [3] es la concentración en sangre de un medicamento cuando se administra una inyección por vía intravenosa en una sola dosis

rápida (en contraposición a la perfusión continua mediante gotero). La distribución del fármaco en los tejidos está afectada por muchos factores y la metabolización se realiza a través del hígado y la orina, entre otras vías. Pero el resultado global es que el organismo elimina la sustancia a un determinado ritmo de "aclaramiento", medido en volumen de plasma o sangre por unidad de tiempo, y la cantidad total de fármaco que permanece en el cuerpo depende de la concentración, que a su vez va cambiando. De este modo se puede predecir la concentración plasmática que habrá en un determinado momento, o bien estimar la dosis que se inyectó inicialmente, o la vía por la que se está excretando, ya que cada sustancia tiene una semivida diferente.

Por ejemplo, un antibiótico como la amoxicilina tiene una semivida de 1-2 horas mientras que la del diazepam (usado para los trastornos de ansiedad) es de 30-60 horas. Conocido este parámetro, habrá que tener en cuenta si el medicamento no es de mucha toxicidad, para dar más espaciadamente dosis más grandes, o si debe fraccionarse a lo largo del tiempo, ya que se sigue acumulando al dar dosis repetidas. Para la mayor parte de los medicamentos hay una concentración por debajo de la cual el fármaco es inefectivo, y otra por encima de la cual es peligroso.

En la perfusión continua, en cambio, se trata de alcanzar una concentración de equilibrio que se mantenga en el tiempo. En la práctica clínica, se considera que la concentración se estabiliza indefinidamente cuando han pasado aproximadamente siete semividas.

Modelos parecidos se usan en otros tipos de administración de cualquier sustancia. Así, la semivida de la cafeína ingerida por una persona sana está entre 1,5 y 9,5 horas, dependiendo de las circunstancias fisiológicas y ambientales (uso de anticonceptivos, fumar, embarazo) [4].

En resumen, desde que E. Rutherford (premio Nobel de Química) formuló por primera vez hacia 1902 la idea de la semivida o "período de semidesintegración" [5] de un isótopo, esta noción se ha convertido en algo común en muchas ramas de la Ciencia. La fórmula exacta que se utiliza en estas situaciones se conoce con el nombre de "decaimiento exponencial", ya que la variable tiempo aparece en el exponente, y es un ejemplo más de cómo se pueden explicar y simular matemáticamente muchísimos fenómenos experimentales, algunos con aplicaciones imprevisibles, en lo que se ha dado en llamar "la irrazonable eficacia de las Matemáticas" [6].

Una última curiosidad: el volumen de la espuma de una cerveza también decae exponencialmente con el tiempo, y la semivida depende del tipo de cerveza. Este descubrimiento le valió a sus autores un premio "Ig Nobel" de Física en el año 2002 [7].

Referencias:
[1] https://es.wikipedia.org/wiki/Anexo:Ganadores_del_Premio_Nobel_de_Qu%C3%ADmica
[2] https://es.wikipedia.org/wiki/Generador_termoel%C3%A9ctrico_de_radiois%C3%B3topos
[3] https://es.wikipedia.org/wiki/Farmacocin%C3%A9tica
[4] Pharmacology of Caffeine http://www.ncbi.nlm.nih.gov/books/NBK223808/
[5] https://es.wikipedia.org/wiki/Periodo_de_semidesintegraci%C3%B3n
[6]https://en.wikipedia.org/wiki/The_Unreasonable_Effectiveness_of_Mathematics_in_the_Natural_Sciences
[7] https://es.wikipedia.org/wiki/Anexo:Galardonados_con_el_premio_Ig_Nobel

Enrique Macías-Virgós
Doctor en Matemáticas
Profesor Titular Universidad de Santiago de Compostela

Capítulo 46
Quarks, leptones y sus interacciones: ¿cuáles son los componentes fundamentales de la materia?

(Por Angel M. Uranga Urteaga)

La sencilla pregunta "¿de qué están hechas las cosas?" ha constituido, desde los primeros filósofos griegos, uno de los motores principales del pensamiento racional de la Humanidad. Hoy día, la respuesta científica a esta pregunta nos sumerge en la física de partículas, que explora las leyes cuánticas de la naturaleza a las escala de longitud más pequeñas accesibles experimentalmente, aproximadamente 10^{-18} metros.

A nuestra escala cotidiana, de metros o centímetros, las sustancias materiales que nos rodean parecen un continuo. Pero descendiendo a escalas del Amstrong (10^{-10} metros), se observa que los objetos están compuestos de moléculas, estados ligados de átomos. Un resultado fascinante de la ciencia del siglo XIX es la comprensión de que toda la materia conocida se reduce a combinaciones de aproximadamente cien tipos de objeto, los átomos de la tabla de Mendeléyev, cuyas regularidades además codifican las propiedades químicas de los elementos correspondientes. Hoy día sabemos que estos átomos no son elementales, indivisibles, sino que están compuestos de electrones con carga eléctrica negativa, que orbitan alrededor de un núcleo cargado positivamente, muy masivo y de tamaño extremadamente diminuto, unos 10^{-15} metros. Es más, mientras que los electrones sí que parecen elementales (hasta la escala de distancias que ha sido posible comprobar experimentalmente), los núcleos están compuestos de protones y neutrones, con masas aproximadamente iguales, pero con cargas positiva y neutra, respectivamente. A su vez, los protones y neutrones tampoco son elementales, sino que están compuestos de tres partículas, los quarks, que sí parecen ser elementales (hasta donde ha sido posible comprobar experimentalmente). Las propiedades de protones y neutrones son consecuencia de las de sus quarks componentes. Por ejemplo, la carga eléctrica de valor +1 (en unidades adecuadas) de un protón, proviene de que está compuesto de dos quarks de tipo "up" (con carga eléctrica +2/3 cada uno) y un quark de tipo "down" (con carga -1/3); mientras que la carga nula de un neutrón, se sigue de que está compuesto de un quark de tipo "up" y dos quarks de tipo "down". Además de por su carga eléctrica fraccionaria, los quarks son partículas peculiares por otros motivos: no existen como partículas libres, sino que necesariamente están confinadas en estados ligados por la interacción de color, que describiremos más adelante.

Por tanto, la materia ordinaria está compuesta de quarks "up", "down" y electrones. Además, están los neutrinos (concretamente los de tipo electrónico), que son partículas prácticamente sin masa y que interactúan muy poco intensamente con cualquier otra, por lo que no son componentes de los átomos y objetos ordinarios. Sin embargo, están muy relacionados con las partículas anteriores: un quark down puede transformarse en un quark up emitiendo un positrón (antipartícula del electrón) y un neutrino. Este proceso está mediado por la interacción débil, que describiremos más adelante.

Los quarks up y down, el electrón y el neutrino electrónico conforman la denominada primera familia de partículas elementales, que componen toda la materia ordinaria que nos rodea. Sorprendentemente, existen otras dos familias de partículas elementales, con una estructura muy similar: la segunda familia contiene dos quarks, denominados de tipo “charm” y “strange” (con cargas eléctricas +2/3 y -1/3, respectivamente, y con interacciones de color), y dos “leptones”: una partícula cargada negativamente denominada el muón (similar al electrón, pero más pesado), y un neutrino muónico con masa casi nula. La tercera familia contiene dos quarks, de tipo “top” y “bottom”, y dos leptones: la partícula tau, cargada negativamente, y el neutrino tauónico. La estructura de cargas e interacciones de la segunda y tercera familia sigue el mismo patrón que la primera, y solo se distinguen en que las masas de sus partículas son mayores. Esto implica que las partículas de la segunda y tercera familias son inestables y se desintegran en una fracción de segundo en partículas de la primera familia. Por tanto, las partículas de las dos familias adicionales no forman parte de la materia estable: para cuando se unen en un estado ligado, como por ejemplo un “átomo de la segunda familia”, se desintegran en partículas de la primera familia, resultando en un átomo ordinario.

No se conoce ninguna razón fundamental por la que deban existir precisamente tres familias, es un hecho experimental sin explicación teórica conocida. De hecho, sería concebible que existieran más familias, aunque en ese caso algunas de sus propiedades deberían ser distintas: por ejemplo, la masa de los neutrinos correspondientes debería ser extremadamente alta, ya que el estudio de la desintegración del bosón Z (ver más adelante) en el CERN, Ginebra, Suiza, impone cotas muy fuertes al respecto. Este tipo de datos experimentales apoyan la hipótesis más sencilla de que el número de familias de quarks y leptones es exactamente tres.

Las partículas de materia, los quarks y leptones que hemos descrito, se manifiestan como elementales a las escalas accesibles experimentalmente. Esto no quiere decir que no estén compuestas de otras partículas más pequeñas, pero si sucede, debería ocurrir a una escala de distancia menor que la explorada, 10^{-18} metros.

Para entender cómo las partículas forman estructuras y componen los objetos, es necesario describir sus interacciones. Un enorme éxito de la física ha sido reducir todas las fuerzas de la naturaleza a cuatro interacciones fundamentales, que actúan sobre las partículas elementales. Son la gravedad, la fuerza electromagnética, la interacción fuerte (o de color) y la interacción débil. La gravedad actúa sobre todas las partículas y es siempre atractiva; a nivel de partículas elementales es increíblemente menos intensa que las demás interacciones, pero es acumulativa, por lo que domina la dinámica en el mundo macroscópico. El electromagnetismo describe todos los fenómenos eléctricos y magnéticos, incluida la luz, y en concreto es la interacción que permite a los núcleos y electrones formar átomos, y de forma residual a los átomos formar moléculas mediante el enlace químico. La interacción fuerte es la que mantiene a los quarks unidos formando protones y neutrones, y de forma residual une a éstos en los núcleos atómicos; opera solo a escalas comparables al radio de un protón, 10^{-15} metros. La interacción débil es menos intensa que la fuerte o la electromagnética, pero es especial por ser la única interacción que transforma el tipo de partículas que participan en una interacción (como advertimos anteriormente, en la desintegración de un quark down en un up, un positrón y un neutrino); esto es esencial para las reacciones en el corazón de las estrellas en las que núcleos de hidrógeno (protones) se fusionan y forman núcleos de helio

(formados de dos protones y dos neutrones, producidos por la transmutación de protones mediante la interacción débil), emitiendo luz y calor, y , en el caso del Sol, posibilitando la vida en la Tierra.

En el contexto de la física cuántica que rige el comportamiento en esas escalas microscópicas, cada interacción tiene una partícula de fuerza asociada, el “cuanto” del campo correspondiente. La interacción electromagnética tiene asociada una partícula, denominada el fotón, sin masa, que interactúa con las partículas cargadas. La interacción fuerte tiene asociadas ocho partículas denominadas gluones, sin masa, que interactúan con los quarks a través de un nuevo tipo de carga, denominado “color”, que toma tres posibles valores. La interacción de color es casi nula a muy cortas distancias, pero extremadamente intensa a distancias mayores que el radio de un protón, por lo que cualquier partícula con carga de color está confinada en estados ligados: mesones, que son combinaciones de quarks y antiquarks, o bariones (como el protón y el neutrón), que son combinaciones de tres quarks con los tres colores posibles). Por ello, no es posible observar quarks aislados, solo como componentes de estados ligados. Sin embargo, la existencia física de los quarks está completamente establecida, ya que se comportan como partículas esencialmente libres a distancias muy cortas, como se ha comprobado en experimentos de dispersión inelástica profunda (Deep Inelastic Scattering). La interacción débil tiene como mediadoras tres partículas, denominadas bosón Z y bosones W. Estas partículas de interacción son enormemente masivas, del orden de 90 veces la masa de un protón. Esta enorme masa es la razón última de la aparente debilidad de la interacción débil: a bajas energías, la probabilidad cuántica de crear un mediador tan masivo es muy pequeña, lo que resulta en una interacción muy poco intensa. A las altas energías accesibles en los aceleradores de partículas actuales, sin embargo, los bosones Z y W se producen copiosamente, y la interacción débil es tan intensa como la electromagnética (y de hecho, ambas quedan descritas en un marco unificado conocido como interacción electrodébil).

Las masas de los bosones Z y W, y más en general las de las partículas elementales, están conectadas con un nuevo fenómeno, asociado con el bosón de Higgs, el último ingrediente (por el momento) en la física de partículas. En el contexto de la Teoría Especial de la Relatividad, cualquier tipo de masa corresponde a una energía interna asociada a la partícula. Por ejemplo, la masa del protón corresponde en un 99% a la energía de los campos de color creados entre los quarks que lo componen. Sin embargo ¿qué tipo de energía interna podría corresponder a una partícula elemental, no compuesta? La respuesta radica en el campo de Higgs, una magnitud que llena el espacio vacío de forma homogénea, y que se acopla a cada partícula con una intensidad diferente. Este acoplamiento contribuye a la energía interna de la partícula, de modo que se refleja en una inercia ante el cambio de movimiento, una masa. La masa de una partícula elemental es simplemente una medida de su acoplamiento con el campo de Higgs. La hipótesis del campo de Higgs predice la existencia de una nueva partícula, el cuanto del campo de Higgs, que recibe el nombre de bosón de Higgs. Esta partícula fue descubierta en el colisionador LHC del CERN, Ginebra, Suiza, confirmado la hipótesis del campo de Higgs, y su papel en la explicación de las masas de las partículas elementales. Y enseñándonos una muy importante lección de profundas implicaciones: el vacío no es la nada.

Las partículas de materia (las tres familias de quarks y leptones), las partículas de interacción (fotón, gluones, y bosones Z y W), y el bosón de Higgs, constituyen el Modelo Estándar (ME), la teoría que actualmente describe los constituyentes últimos de

la materia y sus interacciones. Se trata de un logro excepcional para la especie humana, una construcción conceptual que explica todos los fenómenos naturales desde las escalas macroscópicas hasta las efervescentes fluctuaciones en el interior de los protones dentro de los núcleos de los átomos.

Sin embargo, existe en la comunidad científica una fuerte sensación de que la naturaleza se reserva nuevos ases en la manga, y que los próximos años pueden revelar nuevos tipos de partículas elementales, e incluso nuevos principios sobre su dinámica, más allá del ME. El estudio del cosmos nos revela que la mayor parte de la materia del universo no corresponde a las partículas del ME, sino a un nuevo tipo de partícula que no emite luz y que constituye un fluido cósmico conocido como materia oscura. Las peculiares propiedades del bosón de Higgs sugieren la necesidad de nueva física a escalas alcanzables por el colisionador LHC en los próximos años, en forma de partículas supersimétricas, o de subestructura interna del Higgs, los quarks y/o los leptones. La pequeñísima, pero no nula, masa de los neutrinos observados parece requerir explicaciones que implican nuevos tipos de neutrinos super-masivos y aún por descubrir. Finalmente, el lector atento habrá observado que no hemos mencionado la hipotética partícula de interacción del campo gravitatorio, el gravitón, que aún no ha sido observado experimentalmente. No obstante, la reciente detección de ondas gravitacionales de origen astrofísico por la colaboración LIGO, así como el estudio de la polarización de tipo B del fondo cósmico de radiación de microondas, prometen una mejor comprensión del campo gravitatorio en el futuro. Estos y muchos otros indicios apuntan a que nos estamos asomando a una nueva frontera en la comprensión de las leyes fundamentales del universo.

Angel M. Uranga Urteaga
Doctor en Física Teórica
Profesor de Investigación, Instituto de Física Teórica CSIC

Capítulo 47
¿Qué es el bosón de Higgs? ¿Por qué ha tenido tanta repercusión su descubrimiento experimental?

(Por Víctor Martín Lozano)

El día 4 de Julio de 2012 Fabiola Gianotti y Joe Incandela, portavoces de las colaboraciones experimentales ATLAS y CMS respectivamente, anunciaron el descubrimiento de una partícula que tenía todas las propiedades para ser el bosón de Higgs. Pero, ¿qué es el bosón de Higgs?

En los años 60 se empezó a formar un modelo dentro del marco de la teoría cuántica de campos para explicar las interacciones fundamentales de las partículas elementales. Dicho modelo es conocido hoy en día como Modelo Estándar (ME) de las interacciones fundamentales. Por aquel entonces no se sabía por qué ciertas partículas tenían masa y otras permanecían sin ella. Para dar explicación a este hecho, los físicos Peter Higgs, François Englert, Robert Brout, Gerald Guralnik, Carl R. Hagen y Tom Kibble formularon de manera independiente un mecanismo mediante el cual solucionaban el problema de la masa. Dicho mecanismo se conoce como mecanismo de Higgs y sería el responsable de otorgar masa a las partículas que componen el ME. Asociado a este campo existiría una partícula que podría ser detectada en aceleradores de partículas, hablamos por tanto del conocido como bosón de Higgs. Una partícula con características muy similares a las de dicho bosón es lo que fue encontrado en el CERN por las colaboraciones ATLAS y CMS.

47.1 Mecanismo de Higgs.

Debido a los avances en la física en el siglo XX tanto a nivel teórico como a nivel experimental, los físicos de partículas de aquel tiempo vivieron momentos muy emocionantes. Una plétora de partículas y nuevas interacciones aparecieron en los resultados experimentales y la necesidad de una explicación teórica se hizo fundamental para el entendimiento de la Naturaleza. Tras arduo trabajo por parte de los científicos de la época quedó claro que el modelo de partículas elementales seguía las pautas asociadas a ciertas simetrías. Estas simetrías se asociaron a las interacciones fundamentales que eran observadas entre las partículas. Todo ello quedó enmarcado dentro de lo que hoy día se conoce como ME de las interacciones fundamentales. Este modelo contiene seis leptones (entre los que se incluye el bien conocido electrón) y seis quarks (los protones y neutrones son partículas compuestas de quarks). Tanto quarks como leptones entran dentro de lo que llamamos fermiones, que son las partículas que conforman la materia que conocemos. El secreto del ME radica en clasificar todas estas partículas de acuerdo a cómo se comportan con respecto a ciertas simetrías. Estas simetrías están asociadas a las interacciones que existen entre las partículas, por tanto si una partícula permanece invariante bajo una transformación de una determinada simetría se debe a que no sentiría el efecto bajo esa interacción. Dichas interacciones

son la interacción fuerte, la débil y la electromagnética, estas dos últimas relacionadas entre sí en una simetría mayor llamada electrodébil. Todas ellas tendrían asociadas diferentes partículas, conocidas como bosones mediadores, que serían las responsables de dichas interacciones. El responsable de la fuerza fuerte es el gluon, una partícula sin masa y que tendría 8 versiones diferentes. Los responsables de la fuerza débil son los bosones cargados $W^{\pm}$ y el bosón neutro Z que tienen un peso de aproximadamente de 80 y 90 veces la masa de un protón, mientras que el responsable del electromagnetismo es el fotón, partícula también sin masa.

Sin embargo, no se explicaba desde la formulación matemática de dichas interacciones la aparición de la masa de las partículas. En los años 60 de manera totalmente independiente tres grupos de investigación formados el primero por Robert Brout y François Englert, el segundo por Peter Higgs y el tercero por Gerald Guralnik, Carl R. Hagen y Tom Kibble, formularon un mecanismo a través del cual las partículas adquirirían masa. Este mecanismo, conocido popularmente como mecanismo de Higgs, consiste en un campo que permea todo el espacio, como ocurre con el campo electromagnético. Así pues, todas las partículas en presencia de dicho campo sentirían una fricción debido a su interacción con él. Cada partícula interaccionaría de una manera diferente con este campo, es decir, sufriría diferente grado de fricción. La masa de cada partícula, por tanto, correspondería a esta fricción y la diferencia de masas entre las partículas dependería de la manera en la cual interaccione con dicho campo.

Matemáticamente hablando, el mecanismo de Higgs consiste en la formulación de un campo, el campo de Higgs, que de manera espontánea rompería una de las simetrías del ME, la simetría electrodébil, provocando así el hecho de originar masa a las partículas. Anterior a dicha ruptura, los términos de masa de las partículas no respetaban dicha simetría, por lo tanto solo cuando ésta está rota, las partículas pueden adquirir masa.

47.2 El bosón de Higgs.

Una vez hablado del mecanismo de Higgs podemos preguntarnos, ¿y el bosón de Higgs dónde aparece aquí? Pues bien, asociado al campo de Higgs aparece una partícula escalar que es producida cuando dicho campo se excita, esta partícula es la conocida como bosón de Higgs. Al igual que las excitaciones del campo electromagnético dan lugar al fotón, cualquier excitación producida en el campo de Higgs daría lugar al bosón de Higgs. Matemáticamente si atendemos a la explicación anterior, el campo de Higgs presenta cuatro grados de libertad. Una vez que rompe la simetría electrodébil tres grados de libertad van a parar a los bosones Z, W^{+} y W^{-} , por lo que quedaría un grado de libertad extra correspondiente a lo que conocemos como bosón de Higgs. De esta peculiaridad se dieron cuenta Brout, Englert y Higgs cuando formularon sus investigaciones acerca del mecanismo de Higgs. Ésta es la razón por la que en 2013 la academia sueca otorgase el premio Nobel de física a Peter Higgs y a François Englert solamente (Brout había fallecido en 2011).

Una vez formulado el mecanismo de Higgs, con la consecuente aparición del bosón de Higgs, el ME de las interacciones fundamentales quedaba prácticamente resuelto. Desde ese momento todos los esfuerzos tanto teóricos como experimentales se enfocaron en descubrir a tan escurridiza partícula. Por la parte teórica las propiedades de dicho bosón fueron estudiadas en relación a sus interacciones con las demás partículas

del ME. Debido a su formulación matemática, uno de los parámetros físicos que definen el campo de Higgs permanece libre, por lo que en principio era imposible predecir qué masa tendría el bosón asociado al campo. Esto suponía un problema a la hora de buscar el bosón de Higgs en los experimentos, pues cualquier rango de masa era posible. Sin embargo, era fácil poder calcular la interacción de dicho bosón con las diferentes partículas elementales una vez que la masa de éste era conocida, por lo tanto los físicos teóricos solo podían dar resultados acerca del bosón en función de la masa de éste. A pesar de parecer una situación poco halagüeña se pudo extraer información muy importante acerca de sus propiedades. Una de las más importantes era que si el bosón de Higgs era producido en un experimento automáticamente se desintegraría en partículas del ME. La probabilidad de desintegración en las diferentes partículas dependería como hemos dicho anteriormente de la masa del bosón, sin embargo, una cosa estaba clara, para descubrirlo habría que mirar los productos de desintegración de éste.

Es notable el hecho de que teorías que van más allá del ME mantienen el mecanismo de Higgs presente para otorgar masa a las partículas, tal y como ocurre por ejemplo en el caso de las teorías supersimétricas. En dichos modelos, el bosón de Higgs puede diferir en las propiedades con respecto a las que aparecen en el ME y muchos estudios se han centrado en éstas. Por ejemplo, un gran adelanto que presentan las teorías supersimétricas con respecto al ME es el hecho de que el rango de masas está acotado de una manera más precisa.

Por la parte experimental la mayoría de aceleradores de partículas se han dedicado a buscar el bosón de Higgs entre otras muchas búsquedas. Los aceleradores LEP y Tevatron buscaron con ahínco dicha partícula consiguiendo poner ciertas cotas al rango de masas donde poder encontrarla. Por tanto, toda la expectación de su descubrimiento se centró en el gran colisionador de hadrones (LHC) que se estaba construyendo en el CERN en el que todos los físicos depositaron sus esperanzas para encontrar el bosón de Higgs, la única partícula que faltaba del conocido ME.

47.3 El descubrimiento del bosón de Higgs.

El LHC es un gran acelerador de partículas que colisiona protones para poder, mediante dichas colisiones, estudiar las diversas interacciones que se dan en el ME. Entre dichas interacciones los estudios teóricos preveían la presencia del bosón de Higgs y tras dos años de recogida de datos a una energía en el centro de masas de 7 y 8 TeV (Teraelectronvoltios) [1], desde 2010 a 2012, los resultados acerca del bosón de Higgs fueron expuestos por los portavoces de las dos colaboraciones experimentales ATLAS y CMS que operan en el LHC. Una partícula que se asemeja al bosón de Higgs fue descubierta con una masa de 125 GeV (aproximadamente 125 veces la masa de un protón). Dicho descubrimiento se realizó a través de los canales de desintegración del bosón de Higgs en partículas del ME como fue predicho por los físicos teóricos. El LHC, dado que es un colisionador de protones, produce una gran cantidad de gluones que serían los principales responsables de producir el bosón de Higgs. Una vez producido como hemos visto anteriormente éste se desintegraría en partículas del ME. En particular fue observado en tres canales, desintegración en dos fotones, en dos bosones Z y en dos bosones W. Dichos modos de desintegración se habían detectado con unas condiciones estadísticas adecuadas que nos asegurasen un descubrimiento [2]. Además el estudio de los diferentes canales de desintegración, no solo aquellos que

involucraban fotones, y bosones W y Z, coincidía con bastante exactitud a lo predicho por el ME para dicha masa.

La repercusión mediática fue muy importante, medios de todo el planeta, y en todas las plataformas de información dieron buena cuenta de ello y la noticia acerca de su descubrimiento dio la vuelta al mundo en poco tiempo. Al año siguiente de este gran acontecimiento, en el 2013, se les concedió el premio Nobel a Peter Higgs y François Englert por su predicción teórica acerca de esta partícula.

La partícula que cerraba el círculo del ME se había descubierto. Por fin, dicho modelo se había completado y la última pieza del puzle encajaba perfectamente con todo lo anterior. Sin embargo, diversas observaciones experimentales y resultados teóricos nos hacen pensar a los físicos que debe haber física más allá del ME y que probablemente el bosón de Higgs pueda estar relacionado con dichos fenómenos. Así pues, el descubrimiento del bosón de Higgs ha dado un vuelco a la física de partículas y abre nuevas puertas a las investigaciones tanto teóricas como experimentales para estudiar las implicaciones que tendría esta partícula en nuevos modelos.

Notas:
[1] El electronvoltio (eV) es una unidad de energía. En física de partículas energía y masa están relacionadas mediante la ecuación $E=mc^2$ y por tanto el eV es usado como unidad de masa. Por ejemplo, un electrón tiene una masa de 500 .000 eV=0,5 MeV y un protón aproximadamente 1.000.000.000 eV=1 GeV. Un teraelectronvoltio (TeV) equivale a 1.000.000.000.000 eV.
[2] En física de partículas uno puede decir que existe un descubrimiento cuando tiene una estadística suficiente para poder estar seguro con una probabilidad del 99,99994%, lo que se conoce como 5 sigmas y que equivale a equivocarse en una sobre 3.500.000.

Víctor Martín Lozano
Doctor en Física
Investigador Postdoctoral en la Universidad de Bonn y en el Centro Bethe de Física Teórica.

Capítulo 48

Materia y energía: En una bomba nuclear se convierte materia en energía. ¿Es posible hacer lo contrario y convertir energía en materia?

(Por Angel M. Uranga Urteaga)

En la física tradicional, newtoniana, materia y energía son dos conceptos esenciales, pero muy diferentes. La materia corresponde a los objetos físicos, cuyo movimiento se intenta describir cuantitativamente, y que están caracterizados por su masa, es decir su inercia frente a los cambios de estado de movimiento. Por otro lado, la energía es una entidad más abstracta, y que se puede entender como la capacidad de efectuar un trabajo, es decir, de ejercer fuerzas sobre la materia (así, la energía cinética de movimiento de un objeto puede ejercer fuerzas sobre otro mediante un choque; igualmente, la energía potencial de un objeto describe el almacenamiento de esta capacidad, y produce movimiento cuando se libera). Es un principio familiar y cotidiano que tanto la materia (o la masa) como la energía ni se crean ni se destruyen, solo se transforman.

Sin embargo, estos principios aparentemente inamovibles son solo aproximaciones a una realidad más profunda, que se manifiesta claramente solo cuando se intentan describir objetos con velocidades próximas a la de la luz, el reino de la Relatividad Especial. Uno de los resultados más profundos de la Teoría Especial de la Relatividad de Einstein es la relación entre la materia (o la masa) y la energía. En resumen, cualquier forma de energía presente en un sistema contribuye a su masa, entendida como inercia respecto del cambio de su estado de movimiento. Por ejemplo, consideremos una bola de metal hueca, rellena de agua en reposo, y una bola idéntica pero rellena de agua en agitado movimiento. En Relatividad especial, la energía cinética del movimiento del agua en el segundo caso se manifiesta en que la masa de la segunda bola rellena es ligeramente mayor que la del primero (en general, con una diferencia tan enormemente pequeña que es inobservable en experimentos mecánicos con bolas huecas, pero no nula en principio). Igualmente, una bola hueca llena de radiación electromagnética tiene una masa ligerísimamente mayor que una bola hueca vacía, debido a la energía del campo electromagnético en el primer caso. Como ejemplo final, este principio está relacionado con que la existencia de una velocidad máxima (la de la luz c=300.000 km/s) en Relatividad. Cuando un objeto tiene gran velocidad, su energía cinética se puede traducir a un aumento de su masa (inercia), por lo que cuesta mucho acelerarla aún más. Al aproximarse a la velocidad de la luz, la contribución de la energía cinética a la masa/inercia se aproxima a infinito, por lo que nunca es físicamente posible alcanzar esa velocidad límite.

Centrándonos en objetos en reposo, las ideas anteriores se resumen en la famosa ecuación $E=mc^2$, que relaciona la contribución de una energía (de cualquier tipo) “E” de un sistema con su contribución a la masa “m” de ese sistema. Alternativamente, también establece que una partícula con masa (en reposo) “m” posee, simplemente por el hecho de existir, una cantidad de energía “E”. Esta energía puede imaginarse como energía

potencial, que puede ser liberada cuando el sistema (la partícula) experimenta algún cambio en su masa. Por lo tanto, masa y energía son conceptos intercambiables. Y el principio de conservación de la masa y de la energía como entes independientes no tiene sentido, y es sustituido por el principio de conservación de la masa-energía. Es decir, es perfectamente posible la existencia de procesos en los que cierta cantidad de masa queda liberada en forma de energía, e inversamente, de procesos en los que cierta cantidad de energía se transforma en masa de nuevos objetos o partículas.

Los procesos en los que cierta cantidad de masa de un sistema queda liberada en forma de energía nos resultan familiares, ya que se producen en las reacciones nucleares. Un ejemplo en el contexto cósmico es la fusión de hidrógeno en helio en el corazón de las estrellas, como el Sol. La masa de un núcleo de hidrógeno (un protón) es 1,67 x 10^{-27} kg, mientras que un núcleo de helio (dos protones y dos neutrones unidos) es 6,64 x 10^{-27} kg. El núcleo del Sol es una fábrica que convierte grupos de cuatro protones en núcleos de helio. El proceso por el que dos de estos protones se transmutan en dos neutrones está mediado por la interacción débil, que se trata en otros capítulos; asimismo, en el proceso de fusión se pasan por otros estados intermedios, con núcleos de deuterio, etc. Afortunadamente, solo nos interesa el balance energético entre el estado inicial y el estado final, que es independiente de los pasos intermedios. La diferencia de masa es de aproximadamente 0,04 x 10^{-27} kg, que corresponde a una diferencia de energía de 0,36 x 10^{-11} Julios. Tomando 1 kilogramo de material obtenemos una energía del orden de 10^{15} Julios, aproximadamente 1 Megatón. El Sol quema aproximadamente 6 x 10^{11} kilogramos cada segundo, por lo que emite 10^{11} Megatones por segundo, gracias a los cuales nuestra Tierra es un planeta apto para nuestra existencia.

Un cálculo similar nos permitiría calcular la energía liberada en una bomba nuclear de fusión; o de forma más beneficiosa para la Humanidad, en las futuras centrales nucleares de fusión, cuando este proceso se pueda producir de forma controlada. Otro ejemplo es el de las reacciones de fisión, como las que se producen en las bombas nucleares, o en las centrales nucleares. En este caso, cuando un núcleo de uranio-235 captura un neutrón se descompone en, por ejemplo, un núcleo de bario-141 y uno de kryptón-92, con emisión de tres neutrones. La diferencia de masas entre los sistemas inicial y final se transforma en energía emitida, del orden de 1 Megatón por kilogramo.

Un ejemplo final más próximo a la física de partículas es la aniquilación de partículas de materia con partículas de antimateria. Para cada partícula conocida existe una antipartícula, una partícula con exactamente la misma masa, pero con cargas opuestas. Cuando una partícula y su antipartícula colisionan, se aniquilan y liberan la totalidad de su masa en forma de energía, por ejemplo mediante la emisión de fotones (radiación electromagnética). Introduzcamos, como es habitual, unidades de energía adaptadas a la física de partículas, concretamente 1 GeV = 1,6 x 10^{-10} Julios. El equivalente en energía de la masa de un protón es 0,938 GeV, por lo que una aniquilación de un protón y un antiprotón en reposo liberan una energía de 1,876 GeV, unas 100 veces superior a las reacciones de fusión.

Los ejemplos de reactores o bombas nucleares pueden llevarnos a especular (como ha sido el caso en ciertas obras de ficción) con la posibilidad de construir artefactos de destrucción basados en antimateria. Sin embargo, esto no es factible ya que la antimateria no se encuentra de forma natural en la Tierra (ni en nuestro universo

observable), salvo en efímeros fenómenos como radioactividad natural o rayos cósmicos. Y la producción de antimateria en laboratorios es energéticamente muy costosa, y altamente ineficaz. Por ejemplo, el experimento Decelerador de Antiprotones del CERN, Ginebra, Suiza, produce anti-átomos de hidrógeno; en toda su historia ha conseguido acumular solo 10 nanogramos (es decir, 10^{-8} gramos), cuya aniquilación produciría la muy pequeña cantidad de energía de 10.000 Julios (equivalente por ejemplo al consumo de una bombilla de 60W durante 4 horas).

Hasta ahora hemos descrito procesos en los que la masa se convierte en energía. Pero existen muchos procesos y fenómenos en los que la energía se convierte en masa. Uno de los más cotidianos (y por tanto, desapercibido) radica en la propia masa de un protón. Un protón es una partícula compuesta, hecha de tres quarks (dos quarks de tipo “up” y un quark de tipo “down”). Sin embargo, la masa en reposo de los quarks es muchísimo más pequeña que la masa de un protón, aproximadamente solo un 1%. La masa del protón es en su 99% la manifestación de la energía de los intensísimos campos de la interacción de color (gluones) que mantienen los quarks unidos formando el protón. Es una realización física similar a nuestro anterior ejemplo académico de una bola hueca llena de radiación, con la salvedad de que en este caso el campo de radiación (de color) no es una contribución insignificante (sino la dominante) a la masa.

La física de partículas nos proporciona muchos otros ejemplos de transformación de energía en materia. Los aceleradores de partículas son experimentos basados en este principio. Tomando como ejemplo el LHC, en el CERN, Ginebra, Suiza, los protones (cuya masa en reposo es 0,938 GeV) se aceleran hasta que su energía total es de 6.500 GeV, de modo que en su colisión esta energía puede convertirse en masa de nuevas partículas, posiblemente mucho más pesadas que los propios protones que colisionan. Un ejemplo es el bosón de Higgs, cuya masa es equivalente a aproximadamente 125 GeV, y por tanto sería imposible de producir a partir de dos protones si no fuera por la transformación de energía cinética en materia.

En general, la transformación de energía en materia no es solo una curiosidad académica: es el principio que subyace en todos los experimentos que intentan descubrir nuevas partículas con masas superiores a las observadas hasta el momento, y por tanto el que define la frontera de nuestro conocimiento de lo infinitamente pequeño.

Angel M. Uranga Urteaga
Doctor en Física Teórica
Profesor de Investigación, Instituto de Física Teórica CSIC

Capítulo 49
¿Qué se entiende por dualidad onda-corpúsculo?

(Por Ángel S. Sanz)

49.1 Introducción.

La dicotomía entre onda y corpúsculo, comúnmente denominada dualidad onda-corpúsculo, no es sino una historia que entrelaza dos concepciones o modelos antagónicos a los que se ha recurrido para explicar y comprender la naturaleza física, tanto de la luz, como de la materia. A pesar del intenso debate en torno a ambos modelos a lo largo de más de doscientos años (primero en el caso de la luz y posteriormente, ya en el siglo XX, con la materia), hoy día se puede decir que ambos coexisten sin plantear ningún problema. Se utiliza uno u otro en función de su adecuación al fenómeno o proceso físico investigado y su eficacia en tal investigación, pero no apelando a una firme convicción de que uno sea superior al otro, como sucedió en el pasado.

En términos generales, más allá del ámbito de la Física, la dualidad onda-corpúsculo se puede entender (o, tal vez, deba entenderse) fácilmente como una manifestación más del dualismo, la doctrina o modelo filosófico que considera que son necesarios dos sujetos complementarios esenciales para poder acceder a una comprensión plena de un fenómeno determinado, pero cuya coexistencia simultánea no es posible (en términos metafísicos, se suele decir que ambos sujetos o atributos se encuentran en continua oposición). El dualismo es bastante antiguo, habiendo impregnado diversas escuelas filosóficas a lo largo de la historia, tanto orientales, como occidentales. Un ejemplo paradigmático lo encontramos, sin ir más lejos, en la filosofía griega clásica, en el enfrentamiento entre idealismo platónico y materialismo aristotélico, que unos veinte siglos después, entre los siglos XVI y XVII, darían lugar en Europa al racionalismo y el empirismo, dos corrientes de pensamiento diferentes (aunque complementarias) de entender y acercarse al mundo físico que contribuyeron determinantemente a la explosión de la Revolución Científica.

Volviendo al contexto de la Física y, en particular, lo que denominamos física moderna, es decir, la que surge a partir de las primeras décadas del siglo XX, el dualismo también ha jugado un papel fundamental, y no solo en relación a la dualidad onda-corpúsculo. Por ejemplo, la relatividad especial introdujo el dualismo a través de los pares espacio-tiempo y energía-materia. La teoría nos indica que ambos miembros de cada par son equivalentes y, por tanto, pueden entenderse o explicarse de la misma manera. Sin embargo, más allá de la teoría, lo que observamos en la realidad es que cada uno de ellos se manifiesta de una forma totalmente diferente (no medimos longitudes con relojes ni el tiempo con reglas, por ejemplo). Curiosamente, poco más de una década después de la aparición de la relatividad especial, la relatividad general establecía un nuevo paradigma de dualidad, en el que cada uno de esos pares constituía,

a su vez, el elemento de un nuevo par. Conforme a esta visión, el espacio-tiempo constituye el soporte que gobierna la dinámica de la materia-energía, cuya presencia determina la estructura geométrica del espacio-tiempo.

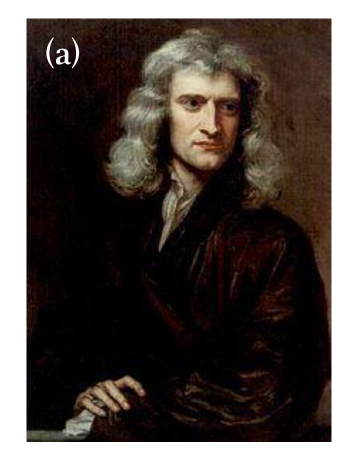
Para tener una cierta perspectiva sobre la idea que subyace a la noción de dualidad onda-corpúsculo es interesante recurrir a lo que denominamos física clásica, es decir, la anterior al siglo XX, que nos sirve para describir el comportamiento de los fenómenos que observamos a nuestro alrededor, como la caída de una manzana o la formación de ondas en un estanque cuando arrojamos una piedra. En el caso de la manzana, es fácil seguir su trayectoria y saber dónde caerá. Esta trayectoria solo variará si se ejerce algún tipo de influencia directa sobre la manzana durante su tránsito (por ejemplo, lanzándole una piedra). En el caso de la onda, la perturbación se extenderá por toda la superficie del estanque, siendo algo más complicado determinar los efectos de la misma, que dependerán de cómo se combinen las diversas ondas formadas. Esto implica que, para determinar el efecto de la perturbación en un instante determinado, sea necesario observar lo que sucede en todos y cada uno de los puntos del estanque. Estos dos fenómenos ejemplifican dos tipos de comportamiento generales que podemos observar en nuestro día a día. Sin embargo, con la aparición de la física cuántica en el siglo XX, el panorama cambió radicalmente, ya que, según ésta, las partículas pueden manifestar ambos comportamientos dependiendo del tipo de experimento a que se sometan, sin importar que se trate de radiación (luz) o materia.

En este capítulo se pretende aportar una visión amplia y general (aunque limitada) de por qué hablamos de dualidad onda-corpúsculo, de dónde y cómo surge este concepto y qué significa. Con tal propósito, en la sección 49.2 se presenta una discusión sobre el origen del concepto de dualidad onda-corpúsculo asociado a la luz. En la sección 49.3 se discuten los aspectos que introdujeron este concepto en el caso de la materia (tanto da si se trata de materia ordinaria, como si es antimateria, pues ambas obedecen las mismas leyes dinámicas) y su influencia en la aparición de la mecánica cuántica. A modo de ilustración, y con objeto de introducir el problema de la medida y cómo la interpretación de Copenhague explica que los sistemas cuánticos puedan ser ondas y corpúsculos. En la sección 49.4 se describen brevemente el experimento de la doble rendija y el de elección retardada de Wheeler. Finalmente, para cerrar el capítulo, en la sección 49.5 se introducirán brevemente las ideas básicas de la mecánica bohmiana, una de las varias formulaciones de la mecánica cuántica, que permite ver cómo onda y corpúsculo no son necesariamente excluyentes (en realidad, tampoco lo son en las formulaciones de Heisenberg y Schrödinger, aunque ello sea menos aparente).

49.2 Luz: rayo, partícula y onda.

Hacia finales del siglo XVII se tenía constancia de diversos fenómenos asociados a la luz. Filósofos chinos, griegos y musulmanes habían estudiado y descrito las sencillas leyes que gobiernan la reflexión y la refracción de la luz, formalizadas durante el siglo XVII por Snell y Descartes, y explicadas por Fermat en términos de un principio general según el cual el tiempo que tarda laluz en viajar desde su fuente hasta su destino es mínimo. También en el siglo XVII Grimaldi había descubierto la propiedad de difracción de la luz, es decir, cuando la luz atraviesa pequeñas aberturas experimenta un efecto similar a doblar una esquina, algo que contravenía la experiencia de siglos de que la luz se propagaba siempre en línea recta, originando sombras bien definidas

(fundamento éste de la cámara obscura, ya conocida varios siglos antes de nuestra era por los filósofos chinos). En 1669, Bartholin describía un fenómeno también asociado a la luz, observado y utilizado como medio de guía en días nublados por los navegantes vikingos, la birrefringencia o doble refracción de la luz. Este fenómeno básicamente consiste en que en materiales anisótropos, como el espato de Islandia (una variedad transparente de la calcita), la luz es refractada a lo largo de dos direcciones diferentes, dando lugar a dos imágenes cuando se observa un mismo objeto (un texto escrito, por ejemplo) con estos materiales. En torno a todos estos descubrimientos tomaron forma hacia finales del siglo XVII y comienzos del XVIII los dos grandes modelos de concebir la naturaleza de luz: la teoría corpuscular de Newton (Fig. 1(a)) y la teoría ondulatoria de Huygens (Fig. 1(b)).





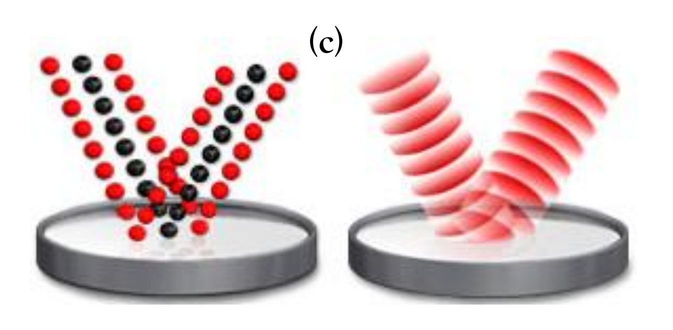


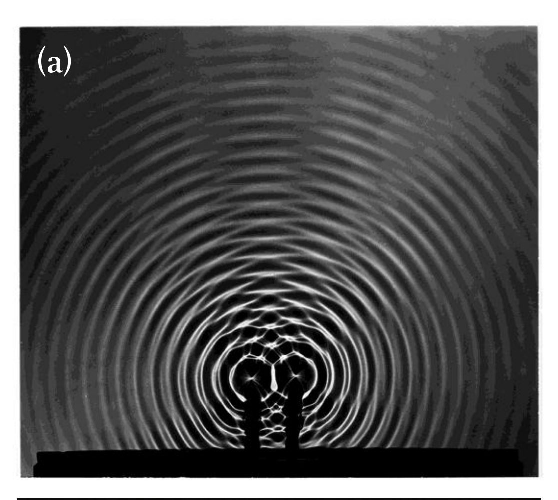
Fig. 1. Isaac Newton (a) y Christian Huygens (b), proponentes de las teorías corpuscular y ondulatoria de la luz. (c) Reflexión de la luz en un espejo plano interpretada con el modelo corpuscular (izquierda) y ondulatorio (derecha).

Uno de los físicos (o, tal vez, por la época, sería más correcto decir filósofos naturales) más destacados e influyentes ha sido Newton, sin lugar a dudas. En sus Philosophiae Naturalis Principia Mathematica [1] condensaba las leyes que rigen el movimiento de los objetos, es decir, las causas o fuerzas que explicaban cómo los cuerpos se desplazan en el espacio por la acción de aquéllas, entre las que se encuentra la gravitación o fuerza gravitatoria. Para ello, se basó en la utilización del concepto de punto material, sin volumen, pero con una masa determinada, lo que permitía simplificar el análisis de los problemas investigados. Una metodología similar fue empleada en su Opticks [2], donde apelaba a la idea de que la luz está compuesta por un gran número de cuerpos diminutos o *corpúsculos* (ver Fig. 1(c), izquierda), que se propagan a gran velocidad y en línea recta a través de un mismo medio. Asignándole una naturaleza corpuscular a la luz, Newton fue capaz de suministró una explicación de sus propiedades, como son el cambio de dirección que experimenta cuando "rebota en" o "se transmite a través de" un medio diferente por el que se estaba desplazando (es decir, la reflexión y la refracción), o cómo se descompone en distintos colores al atravesar un prisma.

Al mismo tiempo, Huygens desarrollaba [3] en Holanda la idea de que la luz, al igual que las ondas de agua que se propagan por un canal al arrojar una piedra, tenía naturaleza ondulatoria. En particular, el modelo de Huygens asumía que cada punto de un frente de ondas era, a su vez, generador de nuevos frentes de onda. A un tiempo dado, el efecto colectivo de todas las ondas generadas por un frente daba lugar a un nuevo frente, lo cual explicaba el avance y propagación de la onda. La idea era muy simple y explicaba bastante bien la propagación rectilínea de la luz, así como los fenómenos de la refracción y la reflexión (ver Fig. 1(c), derecha), algo que también se

conseguía con el modelo corpuscular de Newton. Sin embargo, frente a éste último, era incapaz de explicar el fenómeno de birrefringencia, algo que Newton había conseguido explicar asumiendo que las partículas de luz tenían “lados”. Como consecuencia de ello, así como debido a la gran autoridad científica ganada por Newton en virtud de sus Principia Mathematica, donde no solo establecía las leyes de la dinámica y las leyes del cálculo infinitesimal, sino también su teoría de la gravitación, el modelo de Huygens fue relegado a un segundo plano. La luz era, pues, un conjunto de partículas.

Sin embargo, a diferencia de la teoría de Newton, la de Huygens podía explicar un fenómeno que se observaba con las ondas de agua, pero que aún no se había observado con luz: la interferencia. Supongamos que tenemos dos focos de ondas circulares dentro de una cubeta con agua (ver Fig. 2(a)). Si nos fijamos en un punto cualquiera alejado de los focos, notaremos que hay zonas con una cierta amplitud (crestas) y otras, delimitándolas, en las que no se observa nada, como si no hubiese perturbación alguna. Ello se debe a cómo se combinan sobre cada punto de la superficie del agua los frentes de ondas circularse que proceden de cada foco. Si sobre el punto en cuestión ambos frentes son máximos, observaremos un máximo; si uno es máximo y el otro es mínimo, tendremos una cancelación de la onda. En el caso de la luz, sería en 1803 cuando Thomas Young observaría [4] por primera vez este efecto, que, en términos simples, nos muestra que no siempre luz más luz es igual a luz, sino que también puede resultar obscuridad. Lejos de constituir una violación del principio de conservación de la energía, lo que esto nos indica es que, cuando se trata de ondas, la energía se redistribuye espacialmente de una forma bastante particular, intercalando zonas de luz con zonas de obscuridad (ver Fig. 2(b)). El experimento de Young desbancaba la teoría de la luz de Newton, incapaz de explicar la interferencia. La luz dejaba de ser partícula para convertirse en onda.



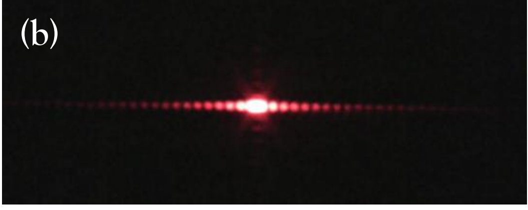


Fig. 2. (a) Interferencia formada en una cubeta de agua mediante dos pulsadores sincronizados. (b) Interferencia formada por el paso de un haz láser rojo a través de dos rendijas.

Aparte del experimento de Young, al asentamiento de la idea de que la luz era onda contribuyó de una forma determinante la generalización del modelo de Huygens que realizó Augustin-Jean Fresnel [5] a principios del siglo XIX, la cual incluía un nuevo fenómeno: la difracción de la luz. Este fenómeno consiste en que, como cualquier onda, la luz puede “doblarse” al llegar a una esquina, de manera que si la hacemos pasar a través de una pequeña abertura, cuanto menor sea ésta, más amplia será la región que abarque la luz tras dicha abertura. Este efecto es el que explica, por ejemplo, que al interponer un objeto con un borde bien definido (un cuchillo afilado, por ejemplo) entre nuestros ojos y una fuente de luz intensa, como puede ser el Sol, se observe una especie de muesca o mella sobre dicho borde, como si la luz fuese capaz de atravesar el objeto (ver Fig. 3(a)). De hecho, aunque no podemos apreciarlo a simple vista, la difracción genera unos patrones característicos (redistribución espacial de la luz) que, en cierto

modo, recuerdan a los de la interferencia, que difuminan la zona que tradicionalmente asignaríamos a lo que sería una sombra bien definida. La importancia de este hecho estriba en que diluye el concepto que desde la Antigüedad se tenía de propagación rectilínea de la luz. Con la difracción ya no se puede hablar de tal cosa, pues entre la región de sombra total y la de luz transmitida existe una zona de transición accesible para la luz (ver Fig. 3(b)).

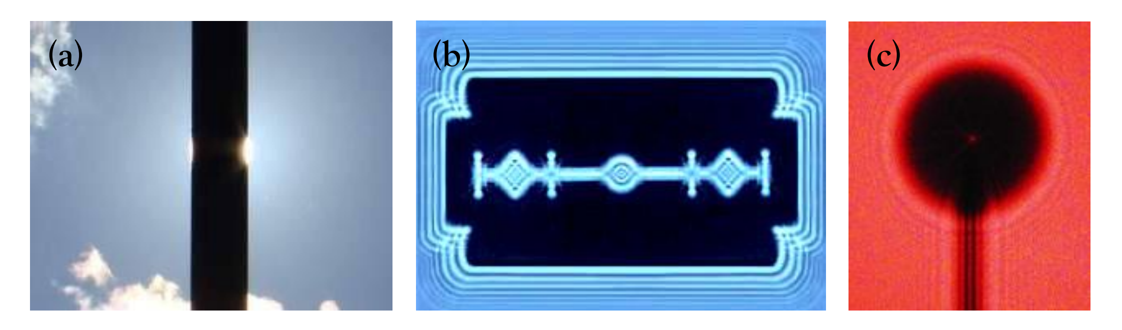
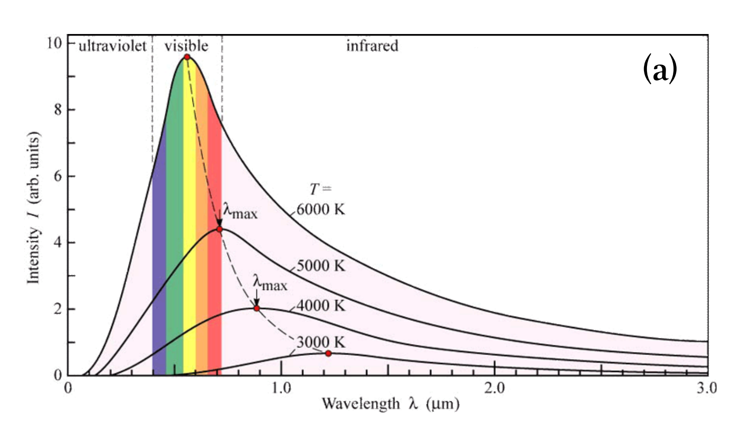


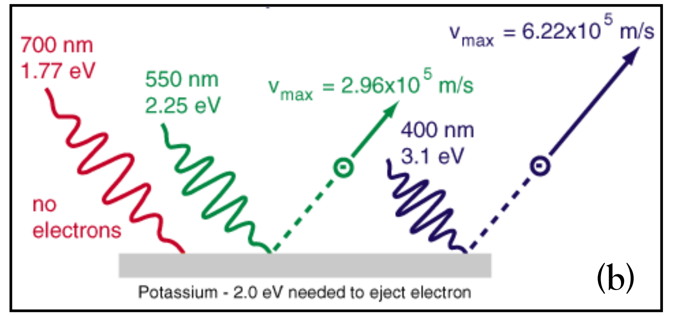
Fig. 3. Diferentes ejemplos de difracción de la luz. (a) Difracción de la luz solar en el borde de un objeto con bordes bien definidos interpuesto entre el Sol y nuestros ojos. (b) Difracción de luz láser azul, donde se aprecia el patrón de franjas más allá de la zona de sombra geométrica. (c) Mancha de Poisson en el centro de la sombra geométrica proyectada por un objeto circular.

Pese al rigor matemático de la teoría de Fresnel, newtonianos convencidos como Siméon Poisson intentaron desbancarla. Según Poisson, a partir de cálculos basados en la teoría de Fresnel, se predecía que justo en el centro de la región de sombra proyectada por un objeto circular debía observarse un punto de luz. Una cosa era observar una región de "penumbra" cerca de los bordes y otra, muy diferente, que ya en una zona de total obscuridad, alejada de los bordes del objeto interpuesto, se observase luz, con independencia del tamaño de dicho objeto. A pesar de la extrañeza del fenómeno, cuando el experimento fue llevado a cabo por François Arago, el resultado fue positivo: efectivamente se observaba la mancha en el centro de la región de sombra (ver Fig. 3(c)), confirmando la teoría de Fresnel y, por tanto, asegurando la supervivencia e implantación del modelo ondulatorio de la luz. En la actualidad a ese punto de luz se le denomina mancha de Poisson o de Arago-Poisson.

Junto con el experimento de Young, el experimento de Arago fue crucial para la supervivencia e implantación del modelo ondulatorio de la luz. A lo largo del siglo XIX, se fue asentando cada vez más, alcanzando su punto álgido con la teoría del electromagnetismo de Maxwell, una teoría que unificaba electricidad y magnetismo bajo la idea de campo elaborada por Faraday a principios de ese mismo siglo (que substituía al concepto más tradicional de fuerza a distancia), y que incorporaba la concepción ondulatoria de la luz de forma natural, como un efecto ligado a la propagación de los campos electromagnéticos. Además, también explicaba de manera natural el fenómeno de polarización de la luz, descubierto por Étienne-Louis Malus en la primera década del siglo XIX, y también el de birrefringencia. Pero no solo esto, la teoría de Maxwell tiene implicaciones más profundas aún, que se irían observando y confirmando a lo largo del siglo XIX, como son la descripción unificada de cualquier tipo de radiación, desde los rayos gamma (la más energética) a las ondas de radio y televisión (la menos energética), pasando por la luz visible, la radiación ultravioleta o las microondas; el valor fijo de la velocidad de la luz, con independencia de la radiación de que se trate; o la invariancia lorentziana de las ecuaciones en las que se basa (descubiertas separadamente por Ampere, Coulomb, Gauss o Faraday), germen de la relatividad especial de Einstein.

Sin embargo, tal como reza la expresión popular "no hay mal que cien años dure, ni cuerpo que lo resista", con el nuevo siglo la concepción de la luz volvería a cambiar. En 1704 Newton publicó en su Opticks que la luz estaba formada por corpúsculos; unos cien años después, en 1803, Young mostraba que la luz era una onda; y otros cien años después, en 1905, Einstein reintroduciría la concepción newtoniana de corpúsculo para explicar el efecto fotoeléctrico. La idea de Einstein surgió de la aplicación de una hipótesis formulada cinco años antes por el físico alemán Max Planck [6], quien para explicar la curva tan característica de un cuerpo negro, es decir, un objeto que puede absorber y emitir radiación de cualquier energía (ver Fig. 4(a)), necesitó considerar una hipótesis crucial para lo que después sería la mecánica cuántica. Frente a la visión tradicional de que a un sistema físico se le puede transferir o éste puede emitir cualquier cantidad de energía para un determinado tipo de radiación (caracterizada por su frecuencia o su longitud de onda), lo que Planck propuso es que esta transferencia solo podía realizarse de forma cuantizada, es decir, mediante un número concreto de "paquetes" indivisibles, cuyo tamaño venía determinado por la frecuencia de la radiación de que se trate (ver ecuación (1)).





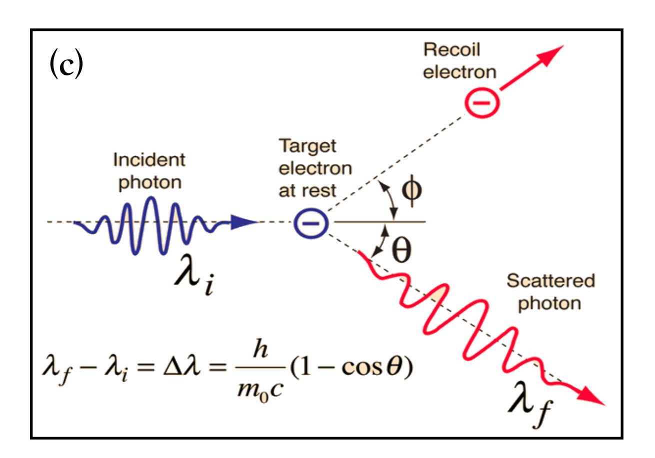


Fig. 4. *(a) Espectro de la radiación emitida por un cuerpo negro para diversas temperaturas. La temperatura superior sería análoga a la superficie del Sol, mientras que la inferior es del orden de la de una lámpara incandescente de filamento de wolframio. (b) Esquema del efecto fotoeléctrico en caso de una lámita de potasio iluminada por radiación de distinta energía. La luz roja es incapaz de extraer electrones, mientras que la verde o la violeta, más energéticas, sí lo son. (c) Esquema del efecto Compton: un fotón con una cierta energía es capaz de desplazar un electrón, como si de una colisión en un billar se tratase. En el proceso, parte de la energía y el momento del fotón son cedidos al electrón.*

En el efecto fotoeléctrico (en el que está basado, por ejemplo, el mecanismo de apertura automática de las puertas de muchos establecimientos), lo que se observa es que, cuando un haz de luz incide sobre un metal, se genera una corriente que no depende de la intensidad de la misma, sino de su frecuencia. Este fenómeno, conocido desde finales del siglo XIX, era inexplicable desde el punto de vista de la teoría electromagnética de Maxwell, como también sucedía con

el espectro de radiación del cuerpo negro. Lo que hizo Einstein fue basarse en la hipótesis de Planck y asumir que el haz de luz estaba compuesto por *quanta* de radiación (corpúsculos a la manera newtoniana). Según este modelo [7], cada *quantum* sería capaz de extraer un electrón si su energía era suficientemente alta, lo cual no depende del número considerado de estos *quanta* (la intensidad del haz incidente), sino de que tuviesen energía suficiente para poder extraer al electrón del metal, lo que equivale a que la frecuencia de la radiación incidente sea mayor que un cierto valor umbral (ver Fig. 4(b)). Lo que observaron Planck y Einstein es que la energía de un *quantum* de radiación viene dada por la expresión

$$E = h\nu = \frac{hc}{\lambda} \quad (1)$$

donde ν y λ son la frecuencia y longitud de onda de la radiación, $h = 6.62607 \times 10^{-34}$ J·s es la constante de Planck y c es la velocidad de la luz en el vacío.

El modelo de Einstein fue confirmado experimentalmente en 1923 por Arthur Compton, quien asumía que si la radiación estaba formada por tales *quanta*, entonces, al igual que cualquier partícula material, estos deberían ser capaces de dispersar otras partículas, como electrones, al impactar sobre ellas (ver Fig. 4(c)). Mediante simples relaciones de conservación de la energía y el momento lineal en la colisión, se obtiene una sencilla relación de dependencia entre la variación de la longitud de onda de la radiación dispersada (fotones dispersados) y el ángulo de dispersión (ver Capítulo 42). El resultado fue un éxito, recibiendo Compton por ello el Premio Nobel en Física en 1927. En cierto modo, este éxito también sirvió para establecer el nombre por el que hoy conocemos al *quantum* de radiación, el fotón, término introducido inicialmente por el químico Gilbert Lewis, en 1926, para denominar a la unidad de energía radiante.

El efecto fotoeléctrico sirvió para establecer la idea de que la radiación (los fotones) posee una naturaleza dual, como onda y como corpúsculo, sin que una deba de prevalecer necesariamente sobre la otra, pues la asignación de un rasgo u otro depende del contexto experimental que se considere (en este efecto se manifiesta la naturaleza corpuscular, de igual manera que en el experimento de Young se observa la ondulatoria). Obviamente, el concepto moderno de corpúsculo trasciende al newtoniano, ya que el fotón, como partícula fundamental, tiene asignadas unas propiedades que van más allá de las meramente geométricas postuladas por Newton. Éstas son las que nos ofrece la teoría cuántica de campos y, a través de ella, el modelo estándar. El fotón es un bosón, sin masa y con espín cero, portador de la interacción electromagnética en virtud del intercambio de un cierto número de los mismos entre dos cualesquiera partículas con carga eléctrica (quarks o leptones). Ello no es impedimento para que, a una energía determinada, pueda manifestar su comportamiento corpuscular, como tal, al que suele apelarse en el efecto fotoeléctrico o en física de partículas (altas energías), o su comportamiento ondulatorio, como ocurre cuando consideramos un gran número de ellos a bajas energías en un experimento típico con rendijas.

49.3 Materia: del punto material a la onda cuántica.

Si bien el debate en torno a la dualidad onda-corpúsculo en el caso de la luz es bastante antiguo, como hemos visto antes, en el caso de la materia es relativamente reciente y está inherentemente ligado a la aparición de la mecánica cuántica a finales del

primer cuarto del siglo XX. No obstante, para comprender bien la dicotomía onda-corpúsculo en el caso de la materia es interesante echar la mirada atrás, a la síntesis de dos corrientes fuertemente asentadas durante los siglos XVII y XVIII, el mecanicismo y el corpuscularianismo, que gozaron de un enorme éxito durante el siglo XIX, pero cuya incapacidad para explicar la estructura interna de la materia dio lugar a la aparición de la mecánica cuántica.

49.3.1 Mecanicismo y localidad.

Esencialmente, el mecanicismo surge como respuesta a la búsqueda de una causa o fuerza que explique el desplazamiento de los objetos, una cuestión tratada por los filósofos griegos, pero que no había experimentado grandes cambios desde la teoría del movimiento de Aristóteles. En los siglos XVI y XVII los filósofos vuelven a plantearse de nuevo estas cuestiones, fundamentalmente a raíz de observaciones del movimiento de los astros y la sistematización de las mismas mediante una serie de leyes relativamente simples. El momento álgido del mecanicismo se alcanza con la publicación en 1687 de los Principia Mathematica de Newton, quien no solo condensa y extiende observaciones y leyes de la mecánica realizadas anteriormente por Copérnico, Kepler, Galileo o Descartes, culminando en sus tres principios de la mecánica y su teoría de la gravitación, sino que también ofrece un método para realizar cálculos dinámicos detallados, el cálculo infinitesimal (paternidad compartida con el matemático y filósofo alemán Gottfried Leibniz). A partir de este trabajo, la complejidad de los objetos se substituye por el concepto de punto material, al que se le asocia la masa del objeto en cuestión y cuyo movimiento viene regido por las leyes de la mecánica de Newton. Este movimiento puede describirse en términos del concepto de trayectoria, una línea imaginaria bien definida que traza el objeto en el espacio durante el curso de su desplazamiento. La modificación de esta trayectoria tendrá lugar si, apelando de nuevo a las leyes de Newton, se ejerce algún tipo de interacción a nivel local sobre el objeto en cuestión.

Las leyes de Newton introducían dos ideas que han permeado la física de los últimos siglos: la causalidad y la localidad. De acuerdo con la causalidad o principio de causalidad, todo efecto observado en un sistema (es decir, en el cambio de su estado) viene originado por una causa que le precede. Conforme a la idea de localidad o principio de localidad, el estado de movimiento de los objetos solo puede ser influido a nivel local, en una vecindad del propio objeto. Aunque ambos conceptos parecen de sentido común, desde la aparición de la física cuántica a comienzos del siglo XX no han dejado de plantearnos retos que aún hoy día siguen generando debate y controversia.

49.3.2 Corpuscularismo y modelos modernos del átomo: de Boyle a Bohr.

La otra corriente a la que hacíamos mención apunta directamente hacia la propia concepción de los objetos materiales, si la materia que los compone es continua y, por tanto, no puede dividirse hasta alcanzar un nivel irreducible en el que pierda sus propiedades, como defendía Aristóteles, o si, por el contrario, está formada por unidades indivisibles e irreducibles, y cuya combinación da lugar a la materia que observamos, como proponía la escuela atomista de Demócrito y Leucipo. Aunque con matices, el atomismo también fue una idea considerada por escuelas indo-budistas antes de nuestra era, así como entre los filósofos del mundo islámico, de los cuales se hicieron eco los filósofos europeos medievales y, posteriormente, los renacentistas, como Descartes, Gassendi, Newton o Boyle. De hecho, es a partir de estos dos últimos, cuando el

atomismo en su versión corpuscularianista (toda la materia está constituida de diminutos corpúsculos) toma mayor impulso, con Newton a través de su teoría corpuscular de la luz y con Boyle a través de su teoría mecánico-corpuscular, cimiento de lo que se convertiría en la (Primera) Revolución Química.

De la teoría corpuscular de la luz de Newton y su influencia en la concepción de la luz durante los siguientes 220 años, hasta la aparición de la mecánica cuántica, ya hemos hablado en la sección anterior. La teoría de Boyle, por su parte, puede verse como el germen del modelo atómico propuesto por John Dalton en 1803, según el cual cada elemento químico (concepto introducido por los filósofos griegos y, en su acepción química moderna, por Boyle) se compone de átomos individuales e inalterables (aún se desconocía la radiactividad y la conversión de isótopos inestables en átomos más estables) de un único tipo, que se combinan para formar estructuras más complejas, los compuestos químicos. Es curioso que el modelo de Dalton se dio a conocer prácticamente al mismo tiempo que Young hacía públicos los resultados de su experimento de interferencia con luz.

Hacia finales del siglo XIX, en 1897, el refinamiento alcanzado por las técnicas de fabricación de bombas de vacío, de construcción de ampollas de vidrio con dispositivos internos, como el tubo de Crookes (o las bombillas de Edison), y la posibilidad de generar altos voltajes mediante bobinas Ruhmkorff (carretes de inducción precursores de los transformadores actuales), permitió observar a Joseph John Thompson que los átomos poseían cargas eléctricas en su interior. Alrededor de una década después, en 1909, Ernest Rutherford demostraba que los átomos no eran una amalgama de electrones y cargas positivas, sino que la carga positiva del átomo se concentraba en una pequeñísima región, en su centro, mientras que los electrones se movían en un volumen mucho mayor, girando en torno a órbitas similares a las descritas por los planetas alrededor del Sol. Este modelo planetario, pese a estar en concordancia con los experimentos y demostrar que los átomos estaban esencialmente huecos, planteaba un problema de consistencia con el electromagnetismo de Maxwell: toda carga (como un electrón) que describe un movimiento circular está emitiendo radiación y, por tanto, en el curso del tiempo perderá energía y acabará cayendo sobre el núcleo del átomo. Obviamente, que los electrones acaben "colapsando" sobre el núcleo del átomo no solo plantea un problema de tipo intelectual, sino también existencial, pues significa que la materia no debería ser estable y, en consecuencia, no hay explicación alguna para la propia existencia de los objetos que observamos a nuestro alrededor, ni tampoco para la nuestra propia, ya que cualquier pedazo de materia (incluidos nosotros) es un conjunto formado por un gran número de átomos. ¿Cómo se explica, pues, la estabilidad de la materia?

En 1913, pocos años después del descubrimiento de Rutherford, Niels Bohr retomaba el modelo de aquél, pero introduciendo una hipótesis crucial: lejos de poder estar a una distancia arbitraria alrededor del núcleo, los electrones se encuentran girando únicamente en ciertas órbitas denominadas *estacionarias*, a una distancia específica del núcleo sin emitir radiación alguna. Solamente cuando el electrón salta de una de estas órbitas a otra se produce la emisión o absorción de un *quantum* de radiación (fotón), cuya energía es equivalente al salto energético entre ambas órbitas. Frente al modelo atómico de Rutherford, el modelo de órbitas estacionarias y saltos cuantizados de Bohr era capaz de explicar las líneas espectrales observadas a lo largo del siglo XIX, desde que Fraunhofer descubriese que el espectro visible de la luz solar, cuando esta se descomponía al atravesar un prisma, no era continuo, del rojo al violeta, como observó

Newton, sino que poseía una multitud de líneas obscuras para las cuales no tenía explicación alguna. Los físicos y químicos del siglo XIX, comenzando por Kirchoff y Bunsen, se dedicaron a la búsqueda de tales líneas y a caracterizar materiales en virtud de las mismas (dos elementos químicos nunca presentan los mismos conjuntos de líneas; son como el DNI de los elementos químicos).

Hacia 1918, el modelo de Bohr había sido refinado con una serie de correcciones realizadas por éste y por Arnold Sommerfeld. Mientras tanto, el físico español Miguel Antonio Catalán había descubierto [8,9] que, a diferencia de los átomos simples, como el hidrógeno, el espectro visible de átomos con una estructura interna compleja estaba formado por grupos de líneas, que denominó multipletes, que surgen por desdoblamiento de las líneas principales en conjuntos de 3, 5 ó hasta 7. Este descubrimiento contribuyó a consolidar el modelo de Bohr y Sommerfeld. Sin embargo, y a pesar de su éxito, los conceptos de órbita estacionaria y salto cuantizado necesitaban aún una explicación satisfactoria. Se desconocía por qué funcionaban y eso suponía un gran obstáculo al modelo en sí, pues no se ajustaba a la teoría conocida hasta el momento.

Obviamente, que la materia estaba compuesta por átomos, unidades individuales e indivisibles con propiedades idénticas para un mismo tipo, como proponía Boyle, era algo que ya se daba por sentado cuando Rutherford, Bohr y Sommerfeld propusieron sus modelos atómicos. Ello se debe, en cierto modo, a las investigaciones sobre una manifestación macroscópica del mundo microscópico: el movimiento browniano. Hacia finales del siglo XVIII, Jan Ingenhousz había observado el zigzagueante e incesante movimiento que exhiben partículas de carbonilla diluidas en agua. Unos cuarenta años después, Robert Brown (de quien deriva la denominación del fenómeno en cuestión) observó [10] el mismo tipo de movimiento en caso de vacuolas de los granos de polen suspendidas en agua. Pese a algunos esfuerzos por intentar comprender la naturaleza de tal movimiento, no fue hasta comienzos del siglo XX cuando se formularon los primeros modelos matemáticos exitosos para describirlo, en particular el de Einstein, de 1905, y el del físico polaco Marian Smoluchowsky, de 1906, que mostraban cómo este efecto constituía una prueba indirecta de la existencia de átomos y moléculas. Las incesantes colisiones de estos últimos, que forman el medio donde se hallan inmersas las partículas de mucho mayor tamaño (visibles al microscopio óptico) que exhiben movimiento browniano, serían la causa de tal efecto. El modelo de Einstein sería verificado por Jean Perrin en 1908, lo que llevó a éste a ganar el Premio Nobel en Física de 1926 ("por su trabajo sobre la estructura discontinua de la materia").

49.3.3 Mecánica cuántica: ondas de materia y no localidad.

Hacia finales del primer cuarto del siglo XX nos encontramos con que los átomos existen y son unidades indivisibles de la materia, con un modelo atómico que es capaz de reproducir las observaciones de que se disponía hasta el momento, pero que carece de toda explicación lógica, y con una teoría clásica (mecánica newtoniana más electromagnetismo maxwelliano) incapaz de explicar ese modelo. La solución a este dilema vino de la mano de Louis de Broglie, quien en 1924, en su tesis doctoral [11], proponía que, si la luz podía comportarse bajo ciertas circunstancias como una partícula, como había mostrado Einstein en su explicación del efecto fotoeléctrico, tal vez la materia ordinaria (los puntos materiales de la teoría newtoniana) podrían exhibir, de algún modo, propiedades ondulatorias. De ser así, si la energía de un *quantum* de un determinado tipo de radiación de frecuencia ν viene dada por la expresión (1), y ésta

debe de ser igual a la expresión relativista para la energía asociada a una partícula sin masa, $E = pc$, tendremos que

$$\lambda = \frac{h}{p} \tag{2}$$

Esta expresión, que relaciona la longitud de onda λ de (la onda asociada a) una partícula con el momento p de ésta, es lo que se conoce en el caso de las partículas materiales como relación de de Broglie. Según esta relación, cuanto más rápida o masiva sea la partícula (recordemos que el momento depende proporcionalmente tanto de la velocidad de la partícula, como de su masa), menor será su longitud de onda y más débil será la manifestación de su naturaleza ondulatoria, y viceversa. Dicho de otra manera, en el límite de longitudes de onda cortas (el límite de la óptica geométrica en el caso de la luz), la materia mostrará su naturaleza corpuscular y su movimiento podrá ser descrito de forma precisa por las leyes de Newton y Maxwell. Esto es lo que sucede con la materia que observamos a nuestro alrededor. Sin embargo, para longitudes de onda largas (partículas más ligeras), las manifestaciones ondulatorias serán más prominentes y tales leyes dejarán de ser válidas.

Ahora, una vez lanzada la hipótesis y establecida la relación (2), queda por ver qué leyes del movimiento sigue esa supuesta onda. Esto fue precisamente lo que hizo Schrödinger [12], quien en 1926 propuso su famosa ecuación de ondas,

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi \tag{3}$$

donde la variable de tipo campo ψ, llamada función de onda, contiene información sobre el estado del sistema, pero sin especificar unívocamente en qué punto del espacio se encuentra y con qué velocidad, como sucede con las leyes clásicas. Físicamente, lo que nos indica la ecuación (3) es que la variación temporal (evolución) del estado de la partícula está directamente relacionada con la distribución espacial instantánea de su energía, que a su vez está determinada tanto por las condiciones de contorno impuestas sobre la onda, como por cualquier tipo de interacción (de tipo interno o externo) que esté actuando sobre la misma. Si tomamos el valor de la función de onda sobre un punto dado del espacio en un determinado instante y calculamos el cuadrado de su módulo (la función de onda ψ es, en general, un campo complejo), lo que obtenemos es la probabilidad de encontrar o detectar partículas en esa región del espacio.

Pese a que pueda parecer un resultado totalmente novedoso, lo interesante de la ecuación (3) es que tiene un punto de partida directamente relacionado con la mecánica clásica: la formulación hamiltoniana. Esencialmente, esta reformulación de la mecánica de Newton, propuesta por William Rowan Hamilton hacia 1833, se basa en substituir el concepto de fuerza a distancia newtoniano por el de energía, lo que permite encontrar una analogía directa entre la dinámica de partículas materiales y la evolución de rayos ópticos (estos serían el análogo directo de las trayectorias newtonianas). De este modo, la evolución del sistema tiene lugar conforme a un mecanismo de minimización de igual forma que en óptica geométrica los rayos proceden de la minimización del tiempo de recorrido de un punto a otro, conforme al principio de Fermat. De hecho, si en vez de energías se considera una magnitud más general denominada acción mecánica (con dimensiones de energía multiplicada por tiempo), la analogía es directa, pues esta

acción es el equivalente de las superficies de fase constante que cualquier rayo cruza perpendicularmente a lo largo de su recorrido. Éste es el punto a partir del cual Schrödinger comenzó la elaboración de su ecuación, generalizando el concepto de fase geométrica, lo que pasaba por incluir además un término de amplitud a la onda relacionado con la intensidad asociada a ésta.

De Broglie propuso la idea de que la materia podía comportarse como una onda y Schrödinger suministró la ecuación matemática que gobierna la evolución de tales ondas. Es por ello que es común utilizar la terminología *onda de materia* para referirse a este comportamiento cuando estudiamos propiedades de difracción o de interferencia con partículas. Este es el caso, por ejemplo, cuando se realiza interferometría con partículas materiales, como neutrones o electrones. Que la materia se comporta como una onda es algo que se ha venido confirmando con electrones, átomos, neutrones, grandes complejos moleculares o nubes de átomos en fase de condensado de Bose-Einstein desde que Schrödinger propusiese su ecuación. Igualmente, apelando a su naturaleza ondulatoria, desde entonces se han venido empleado partículas materiales para estudiar la estructura interna de sólidos y líquidos mediante técnicas basadas en la dispersión de neutrones, electrones o átomos de helio, por ejemplo, que mostraban comportamientos análogos a las técnicas homólogas ópticas (luz visible, ultravioleta o rayos X).

Una característica singular de estas ondas es que, a diferencia del tratamiento clásico del punto material, introducen comportamientos no locales: una perturbación local sobre un punto dado produce un cambio inmediato en la configuración completa de la onda a tiempos posteriores, como sucede en cualquier tipo de onda. En principio, cuando se trata de ondas asociadas a una única partícula, los efectos no son demasiado impactantes. En la sección 49.4 se discute precisamente esta situación mediante dos ejemplos que nos muestran como, lejos de paradojas de gatos encerrados en cajas, lo que realmente sucede con las ondas cuánticas se parece más a lo que sucede cuando se modula una señal mediante el uso de un ecualizador: aunque la señal cambia de forma gradual, la variación de cada frecuencia componente afecta de forma inmediata a toda la señal sin que ello suponga una violación de la relatividad especial (no hay transmisión de información). Más impactante aún es el caso de la onda asociada a dos partículas que han interaccionado en algún momento. De acuerdo con Schrödinger, esa interacción genera una correlación tal que si ambas partículas se separan, las medidas que se realicen sobre ambas seguirán estando altamente correlacionadas con independencia de la distancia que las separe (siempre que no aparezcan otras partículas que puedan interaccionar durante la separación). Es lo que Schrödinger denominó entrelazamiento (cuántico) [13], que contraviene el “sentido común” impuesto por el principio de localidad. Sin embargo, no hay nada incorrecto en ello, pues la mecánica cuántica es inherentemente no local.

49.4 De onda a corpúsculo, y viceversa.

El concepto de dualidad onda-corpúsculo, uno de los más profundos y controvertidos de la física cuántica, es también una de las manifestaciones más palpables del denominado principio de complementariedad [14], enunciado por Bohr en 1927, según el cual la física cuántica prohíbe determinantemente conocer al mismo tiempo dos propiedades o aspectos complementarios de un mismo sistema físico mediante un único experimento, como puede ser la manifestación de su doble

comportamiento como onda y como corpúsculo, dependiendo del experimento que estemos llevando a cabo. La formalización de este principio está directamente relacionado con las relaciones de Heisenberg (ver Capítulo 42). A continuación vamos a ver brevemente dos de los experimentos más paradigmáticos de la física cuántica, que ilustran la dualidad onda-corpúsculo a través de la noción de complementariedad: el experimento de la doble rendija de Young y el experimento de la elección retardada de Wheeler.

49.4.1 El experimento de la doble rendija.

El experimento que Young realizó en 1803 iluminando dos rendijas con luz demostraba que ésta se comporta como una onda al atravesar dichas rendijas, dando lugar a zonas de luz intercaladas con zonas de sombra sobre una pantalla colocada a una cierta distancia de las rendijas (ver panel superior de la Fig. 5). Hoy día, más de doscientos años después de su realización, cuando pensamos en este experimento tan relativamente simple e inocente asumiendo que la luz se propaga por el espacio como las ondas de agua, el resultado nos parece de lo más lógico, podríamos decir. Sin embargo, supongamos que se puede controlar la intensidad de la luz que emerge de la fuente que estemos empleando hasta hacerla tan débil que la energía fuese equivalente, de acuerdo con la relación de Planck-Einstein (1), a un único fotón por unidad de tiempo. Dicho de otro modo, debilitamos tanto la intensidad de la fuente que, al menos en promedio, nos aseguramos de que solamente haya un fotón viajando de la fuente a la pantalla de detección durante el tiempo que tarda dicho fotón en realizar este viaje. Si asumimos la naturaleza corpuscular del fotón, ¿qué sucede cuando éste llega a la doble rendija? Está claro que el patrón de interferencia debe aparecer, tanto si tenemos una fuente de alta intensidad, como si se trata de un único fotón por unidad de tiempo. En este experimento, lo único que la mecánica cuántica nos indica es que, desde el comienzo de su viaje, el fotón ha adquirido una naturaleza ondulatoria, lo que le permite, de algún modo, "desdoblarse" al llegar a las dos rendijas y volver a "recombinarse" posteriormente, contribuyendo al patrón de interferencia que debe observarse.

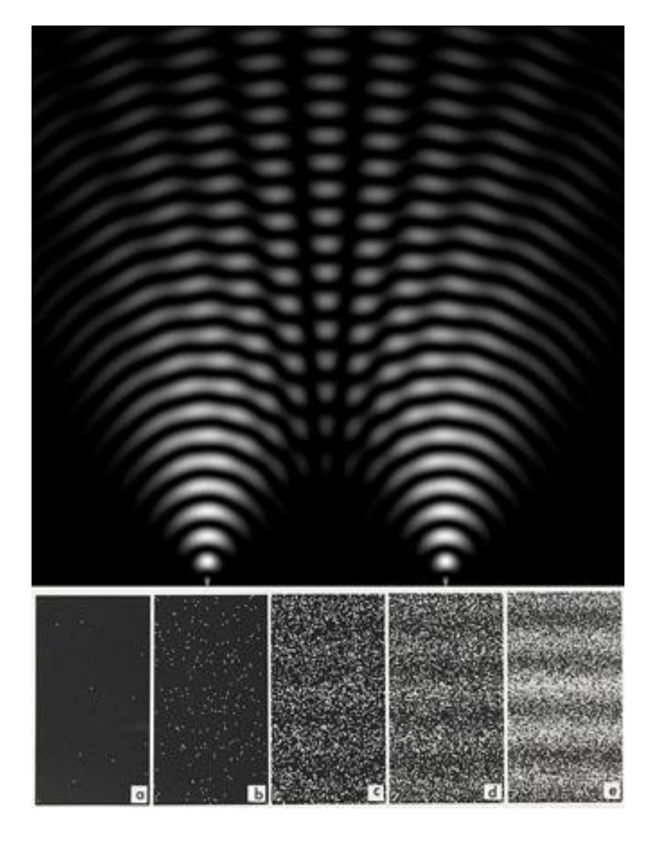


Fig. 5. Panel superior: Recreación de la interferencia de frentes de onda semicirculares emitidos por dos focos sincronizados. Panel inferior: Formación del patrón de interferencia en el experimento de la doble rendija con electrones realizado por Akira Tonomura y colaboradores en 1989 [15]. Se observa cómo a medida que transcurre el tiempo, la acumulación de detecciones de llegadas de electrones pasa de ser aparentemente aleatoria a dibujar una sucesión de bandas claras y obscuras.

Ésa es la visión tradicional y la generalmente aceptada por la forma en que se suele enseñar este experimento. Sin embargo, resulta interesante el hecho de que lo que se observa sobre la pantalla es un impacto y no un diagrama de franjas alternas claras y

obscuras. La mecánica cuántica no dice absolutamente nada sobre cómo ese impacto ha llegado a ese punto de la pantalla. Los fundadores de esta teoría, con el fin de evitar tal problemática, asumieron que tras la evolución del sistema, conforme a la ecuación (3), lo que sucedía era un segundo proceso denominado *colapso de la función de onda*, que tenía lugar por el mero hecho de que la onda era medida. Este colapso sucede de forma totalmente aleatoria, al azar. En nuestro caso, este proceso explicaría que la función de onda asociada al fotón colapse sobre uno de los píxeles de la cámara CCD que dispusiésemos a modo de pantalla de detección (éste es el procedimiento que suele seguirse actualmente en la realización de este tipo de experimentos en el laboratorio; ya no son meramente “experimentos mentales”). Es decir, inicialmente el fotón cambia su naturaleza corpuscular por la ondulatoria por algún motivo totalmente desconocido y, una vez pasadas las rendijas, tiene lugar el proceso inverso durante su detección. Extraño y, en cierto modo, mágico, pero así es como se concibe este experimento. Obviamente, a medida que el número de fotones detectado crece (se repite el mismo proceso un gran número de veces), lo que se obtiene es que las detecciones se distribuyen espacialmente por la pantalle donde son detectadas de tal manera que acaban reproduciendo, por estadística, el patrón de interferencia que se observa con intensidad alta, como se muestra en los sucesivos paneles inferiores (de izquierda a derecha) de la Fig. 5, en el que los puntos blancos son detecciones individuales de electrones.

Pero el experimento no queda ahí. Supongamos que se lleva a cabo una pequeña modificación consistente en poner un detector justo a la salida de una de las rendijas con objeto de poder determinar por dónde pasa el fotón. Esto se puede hacer (y, de hecho, se hace) fácilmente colocando un polarizador lineal sobre cada una de las rendijas de tal manera que los ejes de polarización de cada polarizador formen mutuamente un ángulo recto. De este modo, si el fotón pasa por una de las rendijas, saldrá polarizado en una dirección, y si lo hace por la otra, saldrá polarizado a 90º con respecto a la primera. En este caso, el simple hecho de saber por qué rendija ha pasado el fotón es suficiente como para eliminar el patrón de interferencia que se esperaría y obtener simplemente, tras la acumulación de un número suficientemente grande de fotones, un patrón parecido al que se obtendría con pelotas de golf arrojadas contra una pared con dos orificios, es decir, la suma de lo que sale por cada uno de los orificios. Conforme a la visión tradicional del experimento, lo que sucede es que al tratar de discernir el camino del fotón lo que se provoca es que éste muestre en todo momento su naturaleza corpuscular.

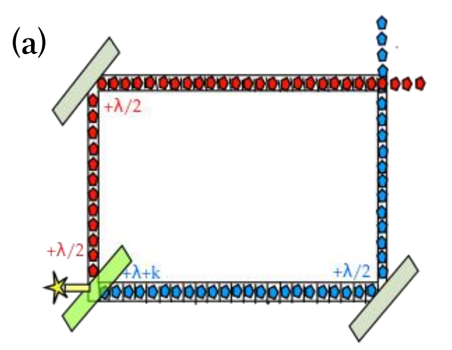
Pero podemos ir un paso más lejos aún. ¿Es posible recuperar el patrón de interferencia? La respuesta es sí. Para ello basta con situar un “borrador cuántico”, es decir, un dispositivo que elimine toda información de por dónde ha pasado el fotón antes de que éste llegue a la rendija de manera que ya no sepamos por dónde pasó el fotón. Dicho dispositivo no es más que otro polarizador lineal colocado de tal manera que su eje esté a 45º con respecto a los ejes de los polarizadores que teníamos sobre cada rendija. Asumimos que entre las rendijas y este nuevo polarizador no hay detector alguno y que, además, el polarizador puede estar a cualquier distancia de las rendijas. Lo que se observa es que, con independencia de por cual de las dos rendijas salga el fotón, éste adquirirá el mismo estado de polarización tras pasar a través del nuevo polarizador, que elimina todo rastro del camino seguido por el fotón y, por tanto, posibilita la formación de un patrón de interferencia en nuestro detector (cuando consideramos un gran número de fotones). El nuevo polarizador ha permitido que el fotón, antes de ser detectado, adquiera por un tiempo su naturaleza ondulatoria.

Este comportamiento tan sumamente extraño no solo se observa con fotones, sino que ha sido verificado una y otra vez con cualquier tipo de partícula material, como electrones, neutrones, átomos, moléculas, grandes complejos moleculares o, incluso, condensados de Bose-Einstein. Actualmente, de hecho se está planteando la posibilidad de realizarlo con objetos incluso mayores, como son cadenas de ácidos nucleicos, virus o bacterias, lo que constituye un verdadero reto (al menos desde la perspectiva más tradicional de la mecánica cuántica) en el sentido de que un organismo vivo podría encontrarse en dos lugares al mismo tiempo (una rendija u otra), igual que en la famosa paradoja del gato de Schrödinger el gato puede estar vivo y muerto, también al mismo tiempo [16].

49.4.2 El experimento de elección retardada de Wheeler.

Hoy día, lejos de ser un experimento mental, el experimento de la doble rendija con fotones individuales (fuentes de luz extremadamente débiles) se puede realizar de forma rutinaria en un laboratorio convencional, pues solo es necesario un láser, una cámara CCD y una serie de dispositivos ópticos, como rendijas, polarizadores y lentes (con partículas materiales es algo más complicado debido al método en que las partículas deben ser preparadas y posteriormente difractadas). El resultado es ciertamente bello a la par que desconcertante. Ahora, la pregunta natural que nos surge ante este experimento es la siguiente: si el fotón debe mostrar un comportamiento determinado, como onda o como corpúsculo en un experimento, ¿qué ocurre si, una vez con el experimento en curso, se produce una variación drástica del mismo? ¿Cambiará el fotón su naturaleza instantáneamente o, por el contrario, seremos capaces de ver ambos aspectos? Esto es precisamente lo que John Archibald Wheeler se preguntaba a finales de 1970 [17].

Lo que planteó Wheeler fue el siguiente experimento (ver Fig. 6). Supongamos que en vez de practicar dos rendijas en una pantalla, lo que hacemos es disponer lo que se conoce como divisor de haz en el camino del fotón, de manera que con una probabilidad del 50% el fotón puede seguir sin desviase (transmitiéndose a través del divisor) o reflejarse 90º con respecto a su trayectoria inicial. En cualquiera de los dos casos, a una cierta distancia se colocan sendos espejos, de manera que cualquiera de las dos posibles trayectorias seguidas por el fotón experimentará un giro brusco de 90º por la reflexión en los espejos. Esto quiere decir que en un cierto instante ambas trayectorias coincidirán sobre el mismo punto (es decir, de nuevo tenemos dos posibles caminos a seguir por el fotón, como en la doble rendija). La sutileza del experimento estriba en lo



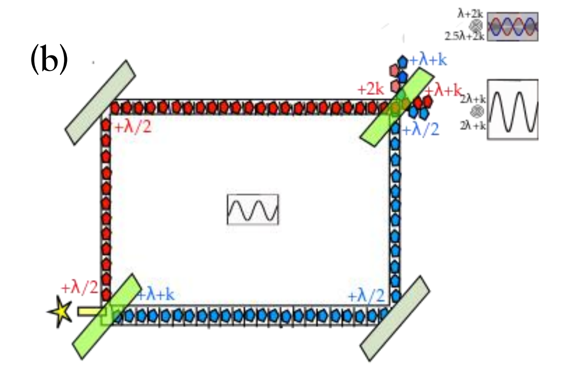


Fig. 6. Experimento de elección retardada de Wheeler con el interferómetro en configuración abierta (a) y cerrada (b). En el segundo caso, variando de forma periódica la diferencia de fase entre ambos caminos es posible modular la señal detectada, cambiando ésta de forma también periódica de un detector al otro.

siguiente. Si a lo largo de la proyección de cada una de las trayectorias tras el punto de corte colocamos un detector (configuración abierta, ver Fig. 6(a)), lo que obtendremos es que el 50% de las veces el fotón será detectado en uno de ellos y el otro 50% en el otro. Si situamos un nuevo divisor de haz sobre el punto de corte (configuración cerrada, ver Fig. 6(b)), el fotón se comporta como una onda (igual que tras el paso por el primer divisor de haz), de manera que las ondas procedentes de ambos caminos se recombinan y solamente se observará señal en el detector situado a lo largo de la prolongación del camino inicialmente reflejado (el 100% de las veces el fotón llegará a este detector).

Lo que nos muestran esos dos experimentos es que, en un caso, pasamos de corpúsculo a onda y de nuevo a corpúsculo; en el otro, la presencia de un nuevo divisor provoca un nuevo paso a onda. ¿Qué sucede si consideramos el primer experimento pero, una vez nos aseguramos de que el fotón ha pasado el primer divisor, modificamos rápidamente el experimento e implementamos el segundo experimento? En principio, tras pasar por el primer divisor, el fotón ha tenido que "decidir" el comportamiento a adoptar y, como no había segundo divisor, será corpúsculo, habiendo así elegido entre un camino o el otro. Sin embargo, el resultado que se obtiene es que solamente uno de los detectores registra señal, conforme a lo que sucede en el segundo experimento. Es decir, a pesar de que el cambio se ha producido una vez el fotón se ha decantado por una de sus dos naturalezas, finalmente "decide cambiar su elección". Igualmente, si inicialmente tenemos colocado el segundo divisor de haz y, una vez asumimos que tenemos al fotón viajando como una onda por ambos caminos, lo quitamos, lo que observamos es señal (del 50%) en ambos detectores. De nuevo, el fotón ha realizado una elección sobre qué aspecto mostrar al final, una elección con cierto retardo, pues no se toma antes de que entre en el dispositivo, sino una vez está dentro y, supuestamente, las "reglas del juego" están ya establecidas.

De nuevo nos encontramos aquí ante un interesante reto cuántico que, cada vez que se realiza en el laboratorio [18], no hace sino corroborar la doble naturaleza de las partículas cuánticas, como onda y como corpúsculo, las cuales, en conformidad con el principio de complementariedad, no nos son accesibles mediante un único experimento, pues o bien observamos una o bien la otra, pero no ambas al mismo tiempo … ¿o, tal vez, esto no es realmente así y lo que hace falta es repensar en otros términos estos experimentos?

49.4.3 Condiciones de contorno: Una mirada pragmática al problema de la dualidad onda-corpúsculo.

En los experimentos que se acaban de describir se ha tenido en cuenta una visión análoga a la de los fundadores de la mecánica cuántica y, particularmente, la que mantenía el propio Bohr con su noción de complementariedad. La cuestión que nos surge, sin embargo, es si realmente es necesario apelar al concepto de corpúsculo. Es decir, si el elemento central de la mecánica cuántica es la onda, ¿cómo se sostiene el concepto de corpúsculo? ¿Es realmente necesario? Ciertamente, contar con un modelo tan sencillo como es el de una masa puntual es bastante atractivo: un objeto que se desplaza por el espacio a una velocidad determinada y cuyo estado de movimiento solo varía si existen fuerzas externas actuando sobre este objeto en un punto dado del espacio. Tratar con ondas, sin embargo, incluso para situaciones relativamente sencillas se torna bastante más complejo, entre otras cosas, porque una onda abarca todo el espacio disponible (accesible para el sistema al que describen). No obstante, una

descripción precisa del efecto fotoeléctrico requiere del uso de ondas, incluso si Einstein consideró un sencillo modelo heurístico que apelaba a la idea de corpúsculo (recordemos que la mecánica cuántica apareció veinte años tras el modelo de Einstein).

Hecha la anterior aclaración, ¿qué es lo que sucede a la partícula cuántica en aquellos casos en que decimos que exhibe su naturaleza corpuscular? Para responder debemos tener en cuenta que toda onda es sensible a las condiciones del contorno que la acota [19] (las ondas en un estanque de agua son sensibles al contorno del propio estanque, así como a la presencia de cualquier otro elemento que pueda encontrarse al nivel de la superficie, como pueden ser rocas o troncos). Si éstas se varían, en el momento que sea, porque el experimento cambia, la onda asociada a la partícula también cambiará como consecuencia de ello, como se indicó al final de la sección 49.3.3, generando un resultado final diferente del que se esperaría sin tal variación del experimento. Así, por ejemplo, en el caso de la doble rendija, lo que está sucediendo cuando situamos un polarizador tras la rendija, no es que repentinamente la partícula deje de ser onda y se convierta, espontáneamente, en un corpúsculo, sino que las ondas que salen por ambas rendijas son *incoherentes* y, por tanto, no dan lugar a interferencia. Esto no quiere decir que no existan efectos cuánticos (manifestaciones ondulatorias), pues si el experimento se lleva a cabo con la suficiente precisión, lo que se observa es que el patrón que se forma es la suma directa de los patrones de difracción que origina cada una de las rendijas por separado. Y, como hemos visto anteriormente en este capítulo, la difracción es, precisamente, un rasgo típicamente ondulatorio. Lo realmente interesante de este experimento es que para que la partícula pueda generar un patrón de interferencia es imprescindible que la distancia entre las dos rendijas sea comparable a la longitud de onda que caracteriza a la onda asociada a la partícula; si esa distancia es mucho mayor, no habrá interferencia alguna, porque la partícula necesita "percibir" que existen dos rendijas abiertas al mismo tiempo.

En el caso del experimento de Wheeler, la explicación es análoga [19]: no es que la modificación del experimento durante su transcurso (y una vez que la partícula está dentro del interferómetro) haga rectificar al fotón su comportamiento, de onda a corpúsculo, y viceversa, sino que, al cambiar las condiciones de contorno del propio experimento, se está influyendo sobre la onda total y, por tanto, el resultado que debe salir. Es decir, la onda sigue siendo onda y lo único que cambia es su reconfiguración (redistribución espacial de la probabilidad asociada) dependiendo de qué es lo que sucede dentro del interferómetro, lo cual no tiene nada que ver con el efecto ilusorio de que el fotón realiza una elección con cierto retardo (algo que, de hecho, inquietaba al propio Wheeler).

Aparte de la variación de las condiciones de contorno, también podrían variar las condiciones físicas a las que está sometida la partícula, como puede ser la interacción con otras partículas o sistemas circundantes. Esto, como apuntaba Schrödinger, da lugar al entrelazamiento irreversible ente los diversos agentes involucrados y, como efecto, tiene lugar lo que se conoce como decoherencia [20,21], es decir, a la pérdida de la coherencia cuántica del sistema por su interacción con otros sistemas, que se traduce en la pérdida de su capacidad de exhibir los rasgos ondulatorios que vendría mostrando de no haberse producido la interacción. Este mecanismo, propuesto por Hans Dieter Zeh en 1970 [22], permite comprender por qué observamos que ciertos sistemas cuánticos se comportan como si estuviesen descritos por las leyes clásicas sin necesidad de abandonar la mecánica cuántica, como mantienen el principio de complementariedad de Bohr y la noción de dualidad onda-corpúsculo.

49.5 Mecánica bohmiana: onda y corpúsculo.

La 5ª Conferencia Solvay, en 1927, sobre la interpretación de la mecánica cuántica [23] finalizó con una visión bastante tajante y pragmática sobre la manera de entender los procesos cuánticos [24]: cualquier cuestionamiento sobre el estado de un sistema cuántico dado no tiene sentido físico alguno a menos que se realice una medida para determinar la propiedad de que se trate, en cuyo caso solo podremos obtener una respuesta parcial, pues el estado del sistema "colapsará" sobre alguno de los posibles valores asignados a tal propiedad y no a todos ellos (una visión completa se obtiene únicamente a través del análisis estadístico de un gran número de estos colapsos). Esta visión es lo que generalmente se conoce como interpretación ortodoxa o de Copenhague de la mecánica cuántica, que recoge principalmente la visión de Bohr sobre cómo han de entenderse los fenómenos cuánticos en relación al proceso de medida. Pese a la sensatez que encierra esta interpretación, el hecho en sí de que se niegue cualquier tipo de cuestionamiento previo a la medida ha originado que durante décadas llegase incluso a negarse la existencia propia del sistema previa a su medida (como sucede con el estado del gato de Schrödinger, vivo y muerto al mismo tiempo hasta que se abra la caja), algo que se ha perpetuado hasta la actualidad.

Sin embargo, ¿existe alguna alternativa ante tal situación? Dejando de lado cuestiones de tipo filosófico y centrándonos en una visión también pragmática del asunto, la respuesta es positiva. Para ello hay que retomar brevemente el modelo propuesto por de Broglie. Su principal contribución, la que nos ha llegado hasta hoy, es básicamente la relación (2). A partir de ella se dice que cualquier partícula cuántica tiene una doble naturaleza, como onda y como corpúsculo, y de ahí el concepto de dualidad onda-corpúsculo que hemos tratado en este capítulo. Sin embargo, y aunque se ha introducido así también en este capítulo, ésa es una idea que surge tras las críticas que el propio de Broglie recibió precisamente en la Conferencia Solvay de 1927. La idea inicial con la que llegó a esta conferencia era algo diferente y consistía en la hipótesis de que toda partícula cuántica era una especie de singularidad que estaba de algún modo ligada irremediablemente a una onda [25,26]. Es decir, había una onda y había también, al mismo tiempo, un corpúsculo; ambas coexistían. La evolución de la onda, cuya longitud de onda era la que aparece en la ecuación (2), determina la dinámica del corpúsculo, cuyo momento es p (por este motivo de Broglie la llamó la onda piloto), pero sin que éste último ejerza, a su vez, ningún tipo de acción sobre la primera. Dicho de otro modo, no es que la partícula cuántica unas veces se comporte como onda y otras como corpúsculo, sino que ambos aspectos coexisten. Fue la incapacidad de de Broglie de suministrar una ecuación de movimiento para la onda lo que determinó en gran medida que su hipótesis original fuese finalmente rechazada.

Como se ha indicado anteriormente, fue Schrödinger quien finalmente encontró la ecuación que regía la evolución de la onda de de Broglie. En 1926, prácticamente tras ser propuesta la ecuación de Schrödinger, Erwin Madelung encontró [27] una sencilla transformación que permitía reescribir dicha ecuación en términos de un par de ecuaciones alternativas análogas a las de la hidrodinámica clásica, hoy denominadas ecuaciones de Madelung. Lo que Madelung buscaba era encontrar una visión menos abstracta de los sistemas cuánticos. Aunque inicialmente no encontró demasiado eco, con el tiempo se ha convertido en una forma estándar de entender la superfluidez, por ejemplo.

A pesar de las propuestas de de Broglie y Madelung, no es hasta 1952 cuando las aguas de la interpretación en mecánica cuántica son fuertemente agitadas, lo que sucede con el modelo planteado por David Bohm [28]. Matemáticamente, este modelo es equivalente al planteado por sus predecesores. Sin embargo, su importancia radica en que replantea por primera vez el problema de la medida y de las variables ocultas desde un punto de vista realmente robusto, no con hipótesis o reformulaciones, sino como un modelo plenamente en pie de igualdad con otros enfoques cuánticos. Sin contravenir ningún resultado fundamental de la mecánica cuántica, Bohm muestra que es posible la coexistencia de unas ondas y de unas trayectorias, donde éstas últimas proceden de la evolución de puntos materiales bajo la guía de las ondas (ondas piloto). La aleatoriedad de la posición inicial (indeterminada e indeterminable) de cada una de esas trayectorias permite escapar a la restricción de la relación de incertidumbre de Heisenberg y, al mismo tiempo, nos permite tener una imagen del proceso de detección (medida) sin necesidad de apelar a la idea de colapso de la función de onda y, por tanto, a la subjetividad de la presencia de un observador externo que nos indique si el gato está vivo o muerto.

Aunque el modelo de Bohm, conocido en la actualidad como mecánica bohmiana (o hidrodinámica cuántica, en la versión de Madelung), fue fuertemente rechazado y olvidado a partir de 1954, en la década de los sesenta John Bell lo retomó, inspirándose en él para formular su teorema y establecer que el problema con el que se encontraron Einstein y otros no era la imposibilidad de encontrar variables ocultas, como tal, sino que la estructura no local de la propia teoría cuántica determinaba qué tipo de variables ocultas podían ser factibles (únicamente las de tipo no local). A partir de los años 70, Bohm y sus colaboradores retomaron estos trabajos [29,30], y a partir de ahí se fue generando una creciente comunidad a nivel internacional con una visión bastante heterodoxa de lo que los sistemas cuánticos son y, por extensión, de cómo debe entenderse el concepto de dualidad onda-corpúsculo.

Hace solamente unos años se realizaron sendos experimentos que mostraban que tal vez esta visión de los procesos cuánticos no solo no es descabellada, sino que posiblemente comience a ser necesaria una revisión en profundidad de lo que realmente se puede decir y no se puede decir en mecánica cuántica, sin mantener el lastre de los tiempos de su fundación, cuando la gran cantidad de experimentos que se realizan hoy día en laboratorios de todo el mundo son una realidad y no una idea. Uno de esos experimentos fue el realizado por el equipo de Aephraim Steinberg en la Universidad de Toronto que mostraba [31] cómo en el caso de la doble rendija de Young es posible determinar los caminos promedio seguidos por los fotones en el transcurso del experimento sin destruir el patrón de interferencia. Otro conjunto de experimentos ciertamente interesantes para el tema que se discute aquí fueron los llevados a cabo independientemente por Yves Couder e Immanuel Fort [32] en la Université Paris Diderot y John Bush [33] en el Massachusetts Institute of Technology (MIT) con fluidos clásicos, en los que se observa cómo una gota (el corpúsculo) en suspensión vibrante sobre una onda en un fluido del mismo tipo es guiada y se comporta de forma similar a como lo hacen sistemas cuánticosanálogos. Es decir, la onda substrato se comporta como la onda piloto de de Broglie. Como no podía ser de otra manera, ambos experimentos tuvieron una amplia repercusión.

Pero, ¿podemos considerar la visión conjunta de onda y corpúsculo que nos suministra la mecánica bohmiana como definitiva? Obviamente, no. La mecánica bohmiana no es más que otra representación adicional de la mecánica cuántica y, por

tanto, en ese sentido, no nos dice nada nuevo que no estuviese ya contenido en ésta. A pesar de ello, resulta interesante, porque nos permite repensar los procesos cuánticos de una manera alternativa a como fueron presentados por Bohr y la escuela de Copenhague. Dicho de otro modo, nos muestra que en mecánica cuántica, al igual que en la mecánica clásica, el formalismo matemático admite una interpretación y comprensión de los fenómenos bastante amplia. Por ejemplo, en el citado experimento realizado por Steinberg y sus colaboradores las trayectorias no indican nada sobre el movimiento particular de cada fotón detectado durante su viaje. Sin embargo, dado que matemáticamente son líneas de flujo, éstas nos permiten visualizar cómo se desarrolla espacialmente, en cada instante, la corriente de fotones, algo que es bastante complicado de visualizar con la representación ondulatoria de Schrödinger o, incluso, imposible si recurrimos a la de matrices de Heisenberg. Por su parte, los experimentos de Couder, Fort y Bush tampoco nos dicen que las partículas cuánticas sean como esa gotita vibrante. Sin embargo, sí que podemos utilizar las trayectorias bohmianas para comprender la dinámica cuántica de la misma manera que la trayectoria trazada por la gotita vibrante nos da una idea del movimiento generado por la onda sobre la que viaja (de la misma manera que una hoja nos permite saber por dónde y cómo fluye la corriente de un río, pero no nos indica nada sobre el incesante movimiento de las moléculas de agua que conforman dicha corriente). En este sentido, el propio Bohm publicó un modelo en 1954 en colaboración con Jean-Pierre Vigier [34], extendido más de tres décadas después con la colaboración de Basil Hiley [35], sobre la naturaleza estocástica del movimiento que subyace a las trayectorias que se deducen del artículo de 1952. En concreto, conforme a ese trabajo (el de 1954), las trayectorias bohmianas no son sino el resultado de promediar conjuntos de trayectorias asociadas a un movimiento difusivo de tipo browniano que evolucionan en un substrato que podríamos denominar "sub-cuántico", es decir, un nivel por debajo del cuántico que describe la ecuación de Schrödinger.

Por el momento, tanto ese tipo de movimiento aleatorio, como el substrato donde se desarrolla, son inaccesibles para la teoría cuántica actual y, por tanto, aunque pueda resultar interesante y atractiva, su hipótesis es solo una mera especulación. Sin embargo, tampoco aquí nos topamos con algo totalmente descabellado. Aunque solemos asumir que la ecuación de Schrödinger es una ecuación de ondas (entre otros motivos, en virtud de la conexión establecida por de Broglie y posteriormente por Schrödinger), lo cierto es que es bastante más parecida a una ecuación de difusión y, en particular, a la ecuación del calor. La ecuación de difusión describe la evolución colectiva de un conjunto de partículas clásicas que se mueven aleatoriamente. A pesar de esa aleatoriedad, las soluciones de la ecuación de difusión son totalmente regulares, como sucede con las funciones de onda de la ecuación de Schrödinger. La principal diferencia con respecto a una ecuación de difusión corriente procede de que en la de Schrödinger el coeficiente de difusión que aparece es una constante imaginaria, $i\hbar/2m$, como ya observó el físico austriaco Harold Fürth en 1933. Esta idea, explorada en mayor detalle en los años sesenta del siglo XX por Kershaw, Comisar o Nelson tomando como base la teoría de procesos estocásticos de Wiener, nos lleva a pensar si, efectivamente, la ecuación de Schrödinger no es en realidad o debe entenderse como una ecuación de tipo estadístico. De hecho, cualquier resultado experimental cuántico no es otra cosa que una estadística realizada con un gran número de sistemas, por hipótesis, no interactuantes y preparados inicialmente con el mismo estado. Dentro de este escenario, las trayectorias bohmianas son el análogo directo a integrar la ecuación del flujo o primera ley de Fick de la mecánica estadística.

Teniendo en cuenta lo anterior, parece que no solo el carácter corpuscular queda totalmente “en el aire”, pues a menos que se apele a éste como mero argumento de discusión (algo que, a noventa años del nacimiento de la mecánica cuántica, y con la tecnología experimental disponible hoy día, parece que no tiene mucho sentido), la teoría en sí no dice absolutamente nada; cualquier comportamiento de ese tipo se explica perfectamente con la mecánica cuántica. Además, vemos que la analogía tan directa entre la ecuación de Schrödinger y la ecuación de difusión nos hace también dudar de si deberíamos hablar aún de una naturaleza “ondulatoria” en vez de simplemente estadística cuántica. Tal vez, al igual que la mecánica clásica acabó desembocando en la termodinámica y la teoría estadística en el siglo XIX, la teoría cuántica habría que mirarla, de forma más correcta, como una generalización de la teoría estadística en general y no solo cuando hablamos de estadísticas de partículas cuánticas (fermiones y bosones). En cualquier caso, y a pesar de ser la teoría más exitosa y potente desarrollada hasta el momento, posiblemente en un futuro, si se somete a mayores pruebas, la mecánica cuántica nos conducirá a resultados y teorías nuevas, como señalaba en una entrevista hace unos años el físico y Premio Nobel en Física Anthony Leggett.

Bibliografía:

[1] “Phisolophiae Naturalis Principia Mathematica”, I. Newton (London, 1687), reimpreso en: “The Principia: Mathematical Principles of Natural Philosophy: A New Translation”, I. B. Cohen and A. Whitman (trads.) (Cambridge University Press, Cambridge, 1999).

[2] “Opticks: Or, a Treatise of the Reflexions, Refractions, Inflexions and Colours of Light”, I. Newton (London, 1704), reimpreso en: “Optiks”, I. Newton (Dover, New York, 1952).

[3] “Treatise on Light”, Ch. Huygens (Leyden, 1690), reimpreso en: “The Wave Theory of Light – Memoirs by Huygens, Young and Fresnel”, H. Crew (American Book Company, New York, 1899), pp. 3-41; también en: “Treatise on Light”, Ch. Huygens (Dover, New York, 1962).

[4] “Experiments and calculations relative to physical optics”, Th. Young, Phil. Trans. Roy. Soc. 94, 1-16 (1804). Reimpreso en: “The Wave Theory of Light – Memoirs by Huygens, Young and Fresnel”, H. Crew (American Book Company, New York, 1899), pp. 68-76; también en: “Great Experiments in Physics”, Sh. Morris (Holt Rinehart and Winston, New York, 1959), pp. 96-101.

[5] “Memoir on the diffraction of light”, A.-J. Fresnel (Paris, 1819), reimpreso en: “The Wave Theory of Light – Memoirs by Huygens, Young and Fresnel”, H. Crew (American Book Company, New York, 1899), pp. 81-144.

[6] “Über eine Verbesserung der Wienschen Spektralgleichung”, M. Planck, *Verh. Dtsch. Phys. Ges. Berlin* **2**, 202-204 (1900), reimpreso en: “The Old Quantum Theory”, D. Ter Haar (Pergamon Press, Oxford, 1967), pp. 79-81 (ver también capítulo 1, pp. 3-14). Ver también: “The Theory of Heat Radiation”, M. Planck (Dover, New York, 1959).

[7] “Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt”, A. Einstein, *Ann. Physik* **17**, 132-148 (1905), reimpreso en: “The Old Quantum Theory”, D. Ter Haar (Pergamon Press, Oxford, 1967), pp. 93-107 (ver también capítulo 1, pp. 3-14).

[8] “Miguel Catalán. Su obra y su mundo”, J. M. Sánchez-Ron (CSIC / Fundación Ramón Menéndez Pidal, Madrid, 1994).

[9] “Miguel Catalán, una obra, un ejemplo”, A. Galindo, *Revista Española de Física* **17**(6), 8-15 (2003).

[10] “A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies”, R. Brown, *Phil. Mag.* **4**, 161-173 (1828).
[11] “Recherches sur la théorie des quanta”, L. de Broglie, Thesis (Paris, 1924).
[12] “An undulatory theory of the mechanics of atoms and molecules”, E. Schrödinger, *Phys. Rev*. **28**, 1049-1070 (1926).
[13] “ Discussion of probability relations between separated systems”, E. Schrödinger, *Math. Proc. Camb. Phil. Soc.* **31**, 555-563 (1935).
[14] “The quantum postulate and the recent development of atomic theory”, N. Bohr, *Nature* **121**, 580-590 (1928).
[15] “Demonstration of single-electron buildup of an interference pattern”, A. Tonomura, J. Endo, T. Matsuda, T. Kawasaki and H. Ezawa, *Am. J. Phys.* **57**, 117-120 (1989).
[16] “Die gegenwärtige Situation in der Quantenmechanik”, E. Schrödinger, *Naturwissenschaften* **23**, 807-849 (1935). Version reimpresa y traducida: “The present situation in quantum mechanics: A translation of Schrödinger’s “cat paradox” paper”, J. D. Trimmer (trad.), *Proc. Am. Phil. Soc.* **124**, 323-328 (1980), reimpreso en: “Quantum Theory and Measurement”, J. A. Wheeler and W. H. Zurek (Eds.) (Princeton University Press, Princeton, NJ, 1983), pp. 152-167.
[17] “The ‘past’ and the ‘delayed-choice double-slit experiment’”, J. A. Wheeler, in “Mathematical Foundations of Quantum Theory”, A. R. Marlow (Ed.) (Academic Press, New York, 1978), pp. 9-48.
[18] “Experimental realization of Wheeler’s delayed-choice GedankenExperiment”, V. Jacques, E. Wu, F. Grosshans, F. Treussart, Ph. Grangier, A. Aspect and J.F. Roch, *Science* **315**, 966-968 (2007).
[19] “Investigating puzzling aspects of the quantum theory by means of its hydrodynamic formulation”, A. S. Sanz, *Found. Phys.* **45**, 1153-1165 (2015).
[20] “Decoherence and the transition from quantum to classical”, W. H. Zurek, *Phys. Today* **44**(10), 36-44 (1991). Versión revisada: “Decoherence and the transition from quantum to classical – REVISITED”, W. H. Zurek, *Los Alamos Science* **27**, 2-25 (2002).
[21] “Decoherence, the measurement problem, and interpretations of quantum mechanics”, M. Schlosshauer, *Rev. Mod. Phys.* **76**, 1267-1305 (2005).
[22] “On the interpretation of measurement in quantum theory”, H. D. Zeh, *Found. Phys*. **1**, 69-76 (1970).
[23] “The Solvay Councils and the Birth of Modern Physics”, P. Marage and G. Wallenborn (Springer, Basilea, 1999).
[24] “Quantum Mechanics – Historical Contingency and the Copenhagen Hegemony”, J. T. Cushing (University of Chicago Press, Chicago, 1994).
[25] “Ondes et mouvements”, L. de Broglie (Gauthier-Villars, Paris, 1926).
[26] “The de Broglie pilot wave theory and the further development of new insights arising out of it”, D. Bohm and B. J. Hiley, *Found. Phys.* **12**, 1001-1016 (1982).
[27] “Quantentheorie in hydrodynamischer Form”, E. Madelung, *Z. Phys.* **40**, 322-326 (1927).
[28] “A suggested interpretation of the quantum theory in terms of ‘hidden variables’ I”, D. Bohm, *Phys. Rev.* **85**, 166-179, 180-193 (1952).
[29] “The Undivided Universe”, D. Bohm and B. J. Hiley (Routledge, London, 1993).
[30] “The Quantum Theory of Motion”, P. R. Holland (Oxford University Press, Oxford, 1993).
[31] “Observing the average trajectories of single photons in a two-slit interferometer”, S. Kocsis, B. Braverman, S. Ravets, M. J. Stevens, R. P. Mirin, L. K. Shalm and A. M. Steinberg, *Science* **332**, 1170-1173 (2011).
[32] “Walking droplets: A form of wave-particle duality at macroscopic scale?”, Y. Couder, A. Boudaoud, S. Protière and E. Fort, *Europhysics News* **41**(1), 14-18 (2010).
[33] “Pilot-wave hydrodynamics”, J. W. M. Bush, *Annu. Rev. Fluid Mech*. **47**, 269-292 (2015).
[34] “Model of the causal interpretation of quantum theory in terms of a fluid with irregular fluctuations”, D. Bohm and J. P. Vigier, *Phys. Rev*. **96**, 208-216 (1954).
[35] “Non-locality and locality in the stochastic interpretation of quantum mechanics”, D. Bohm and B. J. Hiley, *Phys. Rep.* **172**, 93-122 (1989).
[36] “Schrödinger’s kit: Tools that are in two places at once”, M. Brooks, *New Scientist*, Feature 23 de junio, 2010. Enlace: https://www.newscientist.com/article/mg20627661-100-schrodingers-kit-tools-that-are-in-two-places-at-once/

Ángel S. Sanz
Doctor en CC Físicas
Profesor Departamento de Óptica, Facultad de CC Físicas,
Universidad Complutense de Madrid

Capítulo 50
¿Qué son los rayos cósmicos?
(Por Juan Antonio Aguilar Sánchez)

Para explicar los rayos cósmicos hace falta remontarse unos 100 años, al año 1912 cuando Victor Hess finalizaba una serie de viajes en globo aerostático donde equipado con un electroscopio midió cómo la ionización en la atmósfera aumentaba conforme se alejaba de la superficie de la Tierra. El origen de dicha ionización debía ser algún tipo de radiación, y dado que aumentaba con la altura, el origen no podía ser terrestre. En otras palabras, existía, y existe, una radiación proveniente del espacio exterior. Por este hito, a Victor Hess se le conoce como el descubridor de los rayos cósmicos. Sin embargo sería injusto otorgar el mérito solo a Hess dado que muchos físicos antes que él ya habían iniciado el camino que culminaría con sus famosos viajes en globo: Theodor Wulf, Karl Bergwitzy, Domenico Pacini entre otros, fundaron los cimientos de una de las ramas de la física de partículas que dominaría el campo durante los siguientes 40 años hasta el advenimiento de los primero aceleradores de partículas en los inicios de los años 1950.

En los tiempos de su descubrimiento, los rayos cósmicos encerraban numerosos misterios: desde su origen hasta su propia identidad. ¿Qué eran en realidad esos rayos ionizantes? Durante los años 1920 Bruno Rossi y Robert Millikan protagonizaron un animado debate sobre la naturaleza de los rayos cósmicos. Millikan propuso que los rayos cósmicos eran "ultra"-rayos gamma, es decir fotones de muy alta energía creados en la fusión de hidrógeno en el espacio. Las medidas de Rossi, que mostraban una asimetría Este-Oeste en la intensidad de los rayos cósmicos, sugerían en cambio que los rayos cósmicos debían ser partículas con carga eléctrica desmontando las teorías de Millikan. Es famosa la anécdota en la que Rossi, durante la charla introductoria en una conferencia de Roma dijo:

"Claramente Millikan está resentido porque un joven como yo haya hecho pedazos su querida teoría, tanto que desde ese momento se niega a reconocer que existo."

Cien años después sabemos que en efecto Rossi tenía razón (para el descontento de Millikan). En su mayoría, un 90%, los rayos cósmicos son protones y otros núcleos pesados. La proporción de núcleos es tal que sigue fielmente la abundancia atómica que se puede encontrar en nuestro Sistema Solar, lo que apunta a que el origen de estás partículas es estelar. Existen algunas excepciones, por ejemplo, Litio, Berilio y Boro son núcleos que podemos encontrar entre los rayos cósmicos en una proporción mayor de lo que se encuentra en nuestro entorno. Estos núcleos en realidad se producen por la fragmentación de otros más pesados, Carbono en este caso, a lo largo de su viaje por el espacio. Así pues, la relación de abundancias entre Carbono y Boro, nos da información sobre cuanto tiempo el Carbono ha estado viajando por el espacio. El espectro, o

número de partículas por unidad de área y tiempo en función de la energía, también ha sido medido con gran detalle durante los últimos 30 años gracias a la labor de numerosos experimentos. Dicho espectro de rayos cósmicos es sorprendente tanto en su variación como en su rango de energía. El número de partículas, o flujo, cubre 32 órdenes de magnitud, así pues nos encontramos con que las partículas menos energéticas llegan a la Tierra con una frecuencia de una partícula por metro cuadrado cada segundo. Por otro lado, las de más alta energía llegan a un ritmo de una partícula por kilómetro cuadrado, ¡por año! De ahí que los físicos hayan tenido que desarrollar diversas técnicas experimentales para poder medir el espectro de los rayos cósmicos en su totalidad: desde auténticos detectores de partículas enviados al espacio, en satélites o acoplados a la estación espacial internacional, hasta experimentos desplegados en grandes superficie de la Tierra para detectar los rayos cósmicos más energéticos como el Observatorio Pierre Auger que cubre una extensión de unos 3000km^2 en lo alto de la planicie de la Pampa Amarilla, en Malagüe, Argentina.

Pero lo que hace a los rayos cósmicos realmente fascinantes es la cantidad de energía que éstas partículas pueden alcanzar, muy superior a la que se puede conseguir hoy en día con el acelerador más potente construido por el ser humano, el gran colisionador de hadrones (en inglés Large Hadron Collider, LHC) en el Centro Europeo de Investigación Nuclear (CERN) en Ginebra. El LHC es un anillo subterráneo de 27km de longitud, situado en la frontera franco-suiza cerca de Ginebra, Suiza, que usa potentes imanes para acelerar protones al 99.99% de la velocidad de la luz. Pese a lo imponente de este experimento, si tuviéramos que acelerar partículas a las energías de los rayos cósmicos con la misma tecnología necesitaríamos un acelerador del tamaño de la órbita de Mercurio. La velocidad de los rayos cósmicos es tan alta que los efectos de la relatividad especial son realmente considerables. Por ejemplo, pese que el radio de nuestra Galaxia es de unos 100.000 años luz, por la contracción temporal de la relatividad especial los rayos cósmicos más energéticos experimentarían el viaje en tan solo 10 segundos. Cuando llegan a la Tierra los rayos cósmicos se encuentran con 10 kilómetros de atmósfera que, junto al campo magnético de la Tierra, afortunadamente actúan como un escudo y nos protegen de la radiación. Sin embargo, al chocar con los átomos de la atmósfera los rayos cósmicos desencadenan una lluvia de nuevas partículas. Esta lluvia se conoce como rayos cósmicos secundarios y en ellos podemos encontrar una gran fauna de nuevas partículas. Esta es la razón por la que durante muchos años, la física de los rayos cósmicos era el único modo que tenían los físicos de partículas para descubrir y estudiar nuevas partículas. Así pues, siguiendo los pasos de Hess, durante los años 1940 muchos físicos pasaron del laboratorio a globos aerostáticos donde equipados de cámaras de burbujas (una primitiva versión de un detector de partículas) estudiaban esa miríada de nuevas partículas. Entre las nuevas partículas se descubrieron por ejemplo la primera partícula de anti-materia: el positrón, un electrón de carga eléctrica positiva, así como el muon, de propiedades similares al electrón pero de mayor masa.

Pero ¿de dónde vienen los rayos cósmicos? ¿Qué fuente del Universo es capaz de acelerar partículas a tales energías? Esa es la pregunta que, pese a los 100 años desde el descubrimiento de Victor Hess, los físicos aún no hemos sido capaces de resolver por completo. La razón es, sin embargo, fácil de entender. Los rayos cósmicos, al ser partículas con carga eléctrica, son desviados por campos magnéticos durante su viaje por el Universo. Tanto la Vía Láctea como el espacio intergaláctico están inmersos en campo magnéticos, de modo que cuando los rayos cósmicos llegan a la Tierra, su dirección poco o nada tiene que ver con la dirección original lo que imposibilita hacer

astronomía. Sin embargo pese a todo, podemos deducir algunas cosas sobre su origen basándonos, por ejemplo, en su energía. Sabemos que los rayos cósmicos de baja energía deben provenir de nuestra propia Galaxia debido a que los campos magnéticos de la Vía Láctea se ocuparían de confinarlos hasta que éstos acabarían interaccionando con la Tierra. En el otro extremo del espectro, en cambio, los rayos cósmicos de energía extremadamente alta (o UHECR por sus siglas en inglés), deben provenir de fuera de nuestra propia Galaxia, puesto que son tan energéticos que los campos magnéticos de sus respectivas galaxias no serían capaces de retenerlos. El punto de inflexión entre esos dos orígenes es incierto, o en otras palabras, se desconoce cuando exactamente los rayos cósmicos dejan de ser galácticos y pasan a ser extra-galácticos. Sin embargo futuras medidas de precisión del espectro de energía y de la composición de los UHECR nos podrán dar la pista para resolver esa pregunta.

¿Y cuáles serían las fuentes o los objetos encargados de acelerar estas partículas? Lo cierto es que aquí todo son hipótesis dado que jamás se ha observado directamente una fuente de rayos cósmicos. Uno de los objetos candidatos a fuente de rayos cósmicos galácticos son los remanentes de supernova. Al final del ciclo de vida de una estrella ésta puede explorar liberando gran cantidad de masa y energía. Lo que queda detrás puede ser una estrella de neutrones rodeada de todos los restos que han quedado de la estrella original, esto es lo que se llama remanente de supernova (o SNR, otra vez por sus siglas en inglés). Más difícil es imaginar un acelerador cósmico capaz de acelerar partículas hasta la energía equivalente a un balón de fútbol chutado a 50km/h, que son las energías de los UHECR. Aquí la lista de sospechosos se reduce considerablemente dado que existen pocos objetos en el Universo con el campo magnético y el tamaño suficiente para actuar como un gran acelerador de partículas. Los candidatos son los núcleos de galaxia activos y las explosiones de rayos gamma. Los núcleos activos de galaxias son núcleos de galaxias con un agujero negro supermasivo en su interior. Estos núcleos muestran unos haces de partículas en dirección opuesta que podrían funcionar como grandes aceleradores. Por otro lado las explosiones de rayos gamma son los sucesos más violentos conocidos en el Universo y su origen y naturaleza daría para otro capítulo de este libro. De una duración que va desde unos segundos hasta unos minutos, estos eventos son capaces de iluminar todo el cielo liberando su energía principalmente en forma de fotones de muy alta energía.

Pero si los rayos cósmicos nunca apuntan a su fuente, ¿cómo podremos estar jamás seguros de que los núcleos activos o explosiones de rayos gamma son las verdaderas fuentes de rayos cósmicos? Para dar respuesta a este enigma necesitamos de lo que en los últimos años se ha venido llamando la astronomía multi-mensajero. Gracias a la física de partículas sabemos que en las condiciones en las que un protón, por ejemplo, es acelerado a altas energías, se pueden suceder reacciones con la materia de alrededor. Estás interacciones producirían otras partículas como fotones de muy alta energía y neutrinos. Los neutrinos son especialmente interesantes, porque no solo son neutros y por lo tanto viajan en línea recta sin ser desviados por campos magnéticos, sino que son partículas que interactúan débilmente por lo que, al contrario de los fotones, son capaces de atravesar zonas densas del Universo sin ser absorbidos. Identificar varios neutrinos provenientes de una misma zona del cielo sería la prueba inequívoca de que en esa dirección existe una fuente de rayos cósmicos. La dificultad de hacer astronomía de neutrinos deriva precisamente de su virtud: al ser partículas extremadamente elusivas se necesitan detectores de gran tamaño para poder detectarlas.

Hoy en día, el detector de neutrinos más grande del mundo se encuentra en el Polo Sur y consiste en un kilómetro cúbico de hielo Antártico equipado con unos 5.160 fotomultiplicadores. El telescopio IceCube, un nombre que describe muy bien su morfología, es capaz de detectar neutrinos en un gran rango de energías y por tanto identificar neutrinos de varios orígenes. En el año 2013, la colaboración anunció las primeras detecciones de neutrinos de alta energía, compatibles con un origen astrofísico. Sin embargo con apenas un puñado de esos neutrinos aún no es posible identificar las fuentes de rayos cósmicos. Es por ello, que la colaboración ya está pensando en agrandar el telescopio un orden de magnitud. Este nuevo ambicioso proyecto llamado IceCube-Gen2, podrá detectar fuentes 5 veces más débiles que su predecesor y resolver por fin el misterio del origen de los rayos cósmicos.

Juan Antonio Aguilar Sánchez
Doctor en Ciencias Físicas
Investigador, iihe – Université Libre de Bruxelles

Capítulo 51
¿Qué son los neutrinos?
(Por Alberto Casas)

Los neutrinos son partículas fascinantes que ostentan varios «records». Por ejemplo, son las partículas de materia más ligeras que se conocen. Son también, por lo que sabemos, las partículas de materia más abundantes del universo, solo superadas en número por las partículas de luz, los fotones. Y son las partículas conocidas más difíciles de detectar, ya que interaccionan muy poco con la materia, razón por la cual se tardó mucho en descubrirlas, a pesar de su abundancia. Pensemos que los neutrinos atraviesan la Tierra sin inmutarse (y lo mismo harían con muchos millones de Tierras puestas en hilera). Todos estos records pueden servir para otorgar a los neutrinos el título de "partículas más escurridizas". Pero ¿por qué existen los neutrinos y a qué se deben sus enigmáticas propiedades? Antes de profundizar en ello, hagamos un poco de historia.

La existencia de los neutrinos fue predicha por el gran físico austriaco Wolfgang Pauli en 1930. Pauli observó que en la desintegración de ciertos núcleos radiactivos se producía la misteriosa desaparición de una pequeña cantidad de energía, contradiciendo aparentemente el principio de conservación de la energía. Para resolver el problema, Pauli propuso que, en esas desintegraciones, además de los productos de la desintegración visibles, se producía una partícula indetectable y desconocida que se llevaba la energía que faltaba. En 1934 Fermi bautizó la partícula con el nombre de neutrino. Y hubo que esperar hasta 1956 para que fuera detectada por vez primera.

Con el paso de los años hemos aprendido mucho acerca de los neutrinos, aunque aún quedan aspectos esenciales por aclarar. Los neutrinos, que se suelen denotar por la letra griega ν ("nu"), son partículas sin carga eléctrica. Su única interacción conocida, aparte de la gravitatoria, es la llamada interacción débil (ver capítulo 46). Existen tres tipos de neutrinos, llamados neutrino electrónico (ν_e), neutrino muónico (ν_μ) y neutrino tauónico (ν_τ). Estas denominaciones hacen referencia a la partícula cargada (electrón, muón o tau) con la que establecen interacciones. El último neutrino en ser detectado fue el tauónico, ν_τ, en el año 2000; si bien su existencia ya había sido puesta de manifiesto en los años 70. Se podría pensar que puede haber otras especies de neutrinos aún por descubrir, pero no es así. A partir de resultados de aceleradores de partículas se han reunido pruebas convincentes de que no hay más tipos de neutrinos. Mejor dicho, si los hubiera, deberían ser radicalmente distintos de los conocidos.

Los tres tipos de neutrinos son muy ligeros. De hecho, hasta 1998 no había pruebas de que tuvieran masa. Desde entonces sabemos que la tienen, pero no sabemos cuál es, solo ciertos límites entre los que esta tiene que encontrarse. Concretamente, el neutrino más pesado ha de ser entre un millón y cien millones de veces más ligero que el electrón (la siguiente partícula más ligera).

Hablemos un poco de las fuentes de neutrinos en la naturaleza. La mayor parte de los que llegan a la Tierra provienen del Sol, y en cantidades extraordinarias. Ahora mismo usted está siendo atravesado/a, sin notarlo, por cientos de billones de estos neutrinos por segundo. Los neutrinos solares, que no se detectaron hasta 1968, se producen en los procesos de fusión nuclear que tienen lugar en el interior de nuestra estrella. Dichos procesos son complicados, pero en definitiva lo que hacen es convertir protones (o sea núcleos de hidrógeno) en núcleos de helio (formados por dos protones y dos neutrones). Esquemáticamente, por cada cuatro protones se produce un núcleo de helio más dos positrones (electrones positivos), dos neutrinos electrónicos y radiación electromagnética. La radiación producida es la fuente de la luz y el calor que nos llega del Sol. Curiosamente, cada fotón de luz generado tarda cientos de miles de años en salir del Sol, ya que sigue una trayectoria errática al colisionar con las partículas cargadas que encuentra en su camino (protones, núcleos de helio, electrones, ...). Si el horno nuclear de fusión, que es en definitiva el corazón del Sol, se apagara, tardaríamos esos cientos de miles años en notarlo. Sin embargo, los neutrinos escapan instantáneamente del Sol, ya que apenas interaccionan con nada en su camino. Y es gracias a ellos que sabemos que el interior del Sol está a pleno funcionamiento.

Los neutrinos se producen también copiosamente en las explosiones de estrellas (supernovas), y están presentes en los rayos cósmicos que llegan a la Tierra desde el espacio exterior. Por otro lado, son emitidos por todos los núcleos radiactivos que experimentan "desintegración beta" (emisión de un electrón), y se generan asimismo, en cantidades enormes, en las centrales nucleares. Estos últimos escapan de la central nuclear sin problemas pero no son peligrosos, ya que nos atraviesan sin producir el menor efecto. Finalmente, los neutrinos fueron producidos en cantidades extraordinarias al comienzo del universo, en los primeros segundos después del Big Bang. Esos neutrinos primigenios son de hecho los más abundantes en el universo. Sin embargo, son tan poco energéticos que aún no han podido ser detectados. Su descubrimiento es difícil y supondría una gran noticia ya que nos proporcionarían información directa de los primeros instantes tras la gran explosión.

Hasta ahora hemos descrito las propiedades más llamativas de los neutrinos. Pero ¿por qué son así? Para atacar esta cuestión conviene explicar cómo encajan los neutrinos dentro del llamado Modelo Estándar (ME) de la física de partículas, que es el esquema teórico del que disponemos para describir la fenomenología de las partículas elementales (hasta ahora con gran éxito).

En primer lugar, hay que decir que, hasta donde sabemos, los neutrinos son partículas verdaderamente elementales, o sea, no están compuestas de otras partículas. Solo 12 partículas de materia forman este grupo selecto. El protón y neutrón (los habitantes de los núcleos atómicos) no pertenecen a este grupo, ya que están compuestos de otras partículas llamadas quarks. Un protón esta hecho de dos quarks de tipo *u* (up) y un quark de tipo *d* (down). Podemos escribir $p=uud$. De la misma forma, la composición de un neutrón es $n=udd$. Pues bien, el neutrino electrónico, ν_e , junto al electrón, e, y los quarks u y d, forman la llamada *primera familia de partículas elementales* (o familia del electrón):

$$\begin{matrix} \nu_e & u \\ e & d \end{matrix}$$

Existen otras dos familias de partículas elementales, totalmente análogas a la primera: la familia del *muón* (μ) y la del *tau* (τ); de forma que en total tenemos las 12 partículas mencionadas. El muón y el tau son prácticamente idénticos al electrón en todas sus propiedades, excepto en su masa: el muón es unas 200 veces más pesado que el electrón, y el tau unas 3.500. Cada uno tiene su familia completa, formada por un neutrino (ν_μ y ν_τ, respectivamente) y una pareja de quarks (*c*, *s* para el muón, y *t*, *b* para el tau). De hecho, cada familia es una réplica casi exacta de la anterior, excepto que las masas de sus partículas son mayores. Nadie sabe por qué la naturaleza se presenta en esta estructura de tres familias con una jerarquía de masas. El ME no lo explica, sino que parte de este hecho. En ese sentido, el por qué de la existencia de los neutrinos es un misterio, tan grande como la existencia de cualquier otra partícula elemental. Sin embargo, hay que decir que, dada la existencia del electrón y sus interacciones débiles, la estructura matemática del ME *exige* la existencia de una partícula con las características del neutrino electrónico, ν_e. Y lo mismo sucede para las otras dos especies de neutrinos, ν_μ y ν_τ. Así que, en cierto modo, la teoría explica la existencia de los neutrinos y algunas de sus características, como su ausencia de carga eléctrica y sus interacciones débiles. ¿Y su masa? ¿Por qué es tan pequeña?

Desde el punto de vista teórico el origen y la pequeñez de la masa de los neutrinos es quizá la cuestión más interesante acerca de ellos, y está aún sin resolver, aunque hay modelos teóricos muy interesantes que podrían explicarla. De entrada hay que decir que el ME no da ninguna pista sobre las masas de los neutrinos. De hecho, según la formulación original de esta teoría los neutrinos deberían ser exactamente partículas sin masa, debido a la conservación de una cantidad llamada número leptónico (parecida en cierto modo a la conservación de la carga eléctrica). Sin embargo, desde 1998 sabemos que los neutrinos tienen masa, lo que obliga a una reformulación del ME, lo cual es de por sí muy interesante. En realidad, no es difícil extender el esquema teórico del ME de forma que, incluso manteniendo la conservación del número leptónico, las masas de los neutrinos sean distintas de cero. Pero entonces se plantea el problema de por qué son tan excepcionalmente pequeñas. Parece lógico pensar que debe haber alguna razón teórica detrás de esa pequeñez tan extrema.

Una forma alternativa de pensar es suponer que el ME es una teoría efectiva, válida solo hasta una cierta energía, digamos Λ, muy por encima de las energías accesibles en los aceleradores de partículas más potentes. Si la teoría *completa*, que está más allá del ME, viola la conservación del número leptónico, entonces los neutrinos pueden tener masa. Cuanto más grande sea Λ, más pequeña será la masa de los neutrinos (ya que los efectos de dicha violación están más lejanos). Los cálculos indican que si $\Lambda \approx 10^{14}$ veces la masa de un protón, entonces la masa de los neutrinos estaría de forma natural en el rango observado. Si este esquema es correcto, la extrema pequeñez de las masas de los neutrinos nos está informando de física más allá del ME, y sugiriendo algunas de sus características: la no-conservación del número leptónico y la gran escala de energías a la que se encuentra. Para comprobar este esquema, hay que verificar experimentalmente que los neutrinos violan la conservación del número leptónico, algo que se está intentando en experimentos que buscan la denominada desintegración doble-beta de ciertos núcleos atómicos (todavía sin éxito). Así que, si este argumento es correcto, los neutrinos podrían ser los mensajeros privilegiados de la física fundamental que se encuentra a altísimas energías, inaccesible por cualquier otro procedimiento.

Para terminar, vamos a discutir un fenómeno extraordinario, que está en el corazón de todo lo que hemos aprendido sobre los neutrinos en los últimos años: las oscilaciones de neutrinos. Como hemos explicado, existen tres tipos de neutrinos: ν_e, ν_μ y ν_τ. En la jerga de los físicos, estas especies se denominan "sabores" de los neutrinos. Sin embargo los neutrinos físicos, que tienen una masa determinada (aunque aún desconocida), no se corresponden con estos tres sabores, sino que son una mezcla de ellos. Este es un concepto difícil, que entronca con los postulados de la mecánica cuántica, según los cuales una partícula puede estar en una combinación de estados. Imaginemos que tenemos tres botellas llenas de zumo de naranja, de limón y de fresa, respectivamente. Estas botellas representan los tres neutrinos ν_e, ν_μ y ν_τ. Ahora tomamos tres botellas vacías y las llenamos con las tres anteriores pero mezclando los zumos. Estas nuevas botellas, llenas de tres "cócteles" distintos representarían los neutrinos con masa bien determinada, que se suelen denotar ν_1, ν_2 y ν_3. Estos últimos son los que permanecen estables mientras se propagan, mientras que los otros van cambiando de naturaleza (un resultado de la mecánica cuántica). Esto significa que si creamos un neutrino electrónico, ν_e, por ejemplo en el interior del Sol, al cabo de un tiempo existirá una cierta probabilidad de que al detectar dicho neutrino nos encontremos que se ha transformado en ν_μ o ν_τ. Para que este fenómeno se dé, es necesario que los neutrinos tengan masas y que estas sean diferentes entre sí. Este es el fenómeno de oscilaciones de neutrinos que fue observado por vez primera en los neutrinos provenientes del Sol y en los que se crean en la atmósfera como resultado del impacto de rayos cósmicos ("neutrinos atmosféricos"). De esta forma se han podido determinar las diferencias de masa entre los neutrinos (aunque no su masa absoluta) y los llamados "ángulos de mezcla", que determinan como se mezclan los neutrinos (el contenido de los cócteles anteriores).

En definitiva, sabemos mucho acerca de estas partículas singulares, pero aún quedan cosas trascendentales por entender, las cuales podrían darnos pistas cruciales sobre la teoría última de la naturaleza.

Alberto Casas
Doctor en Física
Profesor de Investigación, Instituto de Física Teórica CSIC

Capítulo 52
¿Qué se entiende por materia oscura y energía oscura?

(Por Juan García-Bellido Capdevila)

La naturaleza de la materia y energía oscuras es uno de los mayores misterios de la cosmología moderna. Gracias a la espectacular revolución en las observaciones astronómicas desde hace dos décadas, creemos que conocemos bastante bien el origen y la evolución del universo desde las primeras fracciones de segundo hasta los trece mil ochocientos millones de años del universo actual, así como el contenido de materia ordinaria en forma de galaxias, cúmulos y supercúmulos de galaxias, cuya luz nos llega desde los confines del universo observable. Sin embargo, aún desconocemos de qué está hecho más del 95% de todo el universo, esto es, la materia y la energía oscuras. Lo primero que debemos aclarar es que se trata de dos componentes muy distintas. A pesar de la famosa fórmula de Einstein, no debemos confundir materia oscura con energía oscura. La primera atrae y la segunda repele gravitacionalmente.

52.1 Materia Oscura.

Uno de los problemas más acuciantes y aún sin resolver de la cosmología moderna es el hecho de que el 85% de toda la materia en el Universo esté en la forma de una componente no luminosa, la llamada materia oscura. Propuesta por primera vez en 1933 por Fritz Zwicky para explicar el movimiento de las galaxias alrededor de los cúmulos, su naturaleza ha desconcertado a los astrónomos desde entonces. Hoy en día, tenemos amplia evidencia de la existencia de la materia oscura, procedente de varias observaciones astronómicas, como las curvas de rotación de las galaxias, en las que hace falta más materia que la luminosa para dar cuenta de las velocidades observadas de las estrellas y el gas en dichas galaxias. Una componente importante de materia oscura también es necesaria en la formación de galaxias, para explicar cómo las estructuras a gran escala en el Universo (galaxias, cúmulos y supercúmulos) se originaron a partir del colapso gravitacional de pequeños grumos de gas. Tenemos un conocimiento preciso de la distribución espacial de estos grumos a través de la huella que dejaron en las pequeñas fluctuaciones de temperatura medidas en el fondo cósmico de microondas.

También necesitamos la materia oscura para explicar las desviaciones gravitacionales observadas en la luz procedente de fuentes luminosas distantes detrás de las galaxias y los cúmulos, un fenómeno conocido como lente gravitacional. Todas estas observaciones muestran que la materia oscura interactúa gravitacionalmente con la materia ordinaria, pero no da ninguna pista acerca de que, en realidad, está hecha, y si tiene otras interacciones. Muchas hipótesis diferentes se han formulado sobre la naturaleza de la materia oscura y sus interacciones, con más de noventa órdenes en magnitud en el rango de masa de sus posibles componentes, desde los axiones ultraligeros a los agujeros negros masivos.

La hipótesis mejor explorada considera que la materia oscura está hecha de nuevas partículas masivas de interacción débil (WIMPs por sus siglas en inglés), que se predicen en ciertas extensiones del Modelo Estándar de la física de partículas. Hasta la fecha no hay ninguna evidencia de su existencia, proveniente de búsquedas directas en el colisionador gigante de hadrones (LHC) en el CERN, ni de ninguno de los detectores bajo tierra, como el CDMS o LUX [1]. Por otra parte no se ha encontrado ninguna evidencia indirecta a través de satélites astronómicos como Fermi, en busca de señales de su aniquilación.

Una hipótesis alternativa muy natural, que fue considerada durante mucho tiempo, es la de objetos compactos masivos del halo galáctico (MACHOs), tales como planetas, enanas marrones, estrellas de neutrones o agujeros negros. Tales objetos compactos se buscaron hace décadas, gracias al fenómeno de microlente gravitacional: cuando un objeto masivo, como un agujero negro, una enana marrón o incluso un planeta cruza la línea de visión de una estrella distante, se induce un efecto de lente gravitacional en los rayos de luz de la estrella, lo que resulta en una amplificación temporal acromática de la luminosidad de la estrella.

Algunos experimentos astronómicos buscaron variaciones en el brillo de millones de estrellas en las nubes de Magallanes, sin encontrar ninguna evidencia clara de objetos compactos que comprendan el 100% de nuestro halo galáctico. Estos resultados fueron suficientemente concluyentes como para restringir la hipótesis MACHO hasta masas del orden de unas pocas masas solares. Recientemente, con el descubrimiento por parte de LIGO de ondas gravitacionales provinientes de la colisión de dos agujeros negros de unas 30 masas solares, se ha abierto la posibilidad de que la materia oscura esté formada por una distribución de agujeros negros primordiales de distintas masas, que se formaron en el universo primitivo (antes de la nucleosíntesis de los primeros elementos atómicos), cuando el universo tenía menos de un segundo de vida, a partir de las mismas fluctuaciones que vemos en el fondo de radiación, pero de mayor amplitud a pequeñas escalas.

Si la materia oscura fueran agujeros negros masivos, éstos podrían constituir las semillas a partir de las cuales se formaron las galaxias hace miles de millones de años. La acreción de gas alrededor de dichos agujeros negros formaría los primeros quásares, que aumentarían su masa hasta llegar a los miles de millones de masas solares que se observan en el centro de todas las galaxias espirales. Además, con esta hipótesis se pueden explicar fácilmente muchas propiedades de la distribución de materia oscura en el Universo, y en particular porqué hasta ahora no habíamos detectado las galaxias enanas que orbitan nuestra propia galaxia.

Por otra parte, estos agujeros negros se podrían detectar con GAIA, buscando pequeñas desviaciones en las trayectorias de las estrellas de nuestra galaxia debido a la presencia entre ellas de dichos agujeros negros masivos, así como con experimentos de microlente gravitacional de larga duración que sean sensibles a masas superiores a 10 masas solares. La confirmación definitiva probablemente vendrá de la mano de los detectores de ondas gravitacionales, pues éstos son directamente capaces de detectar la masa y el espín de los agujeros negros que colisionan, no solo hoy en día sino hace miles de millones de años, y por tanto determinar su origen no estelar.

Universe

Dark Energy
Dark Matter
Protons
Helium
Photons
Neutrinos

52.2 Energía Oscura.

No solo la materia oscura es un misterio que podría tener solución en un futuro próximo; también la energía oscura podría tener los días contados, aunque éste es probablemente un problema bastante más peliagudo. Mientras que la materia oscura actúa como la materia ordinaria, atrayendo otras componentes de materia, la energía oscura es completamente distinta: se trata de un agente repulsor del espacio-tiempo, aunque no haya ningún tipo de materia en él. Es una propiedad muy peculiar de la relatividad general y no tiene su análogo newtoniano. En la teoría de Einstein, el espacio-tiempo es dinámico y la materia y la energía lo curvan o deforman. En particular, una densidad de energía aproximadamente constante (como la así llamada constante cosmológica) hace que dos puntos cualesquiera del universo se separen de forma acelerada, cada vez más deprisa. Esta constante fue introducida por Einstein en sus ecuaciones para contrarrestar la atracción gravitacional de la materia, en su búsqueda de una solución estática del universo, antes de la detección de la expansión del universo por Edwin Hubble [2].

A finales del siglo pasado dos grupos confirmaron, usando medidas de la luminosidad de ciertas supernovas lejanas, que el universo se está expandiendo de forma acelerada, tal como predicen las ecuaciones de Einstein con una constante cosmológica. Hoy en día las medidas de dicha aceleración se hacen con una gran variedad de observaciones, desde el fondo de radiación de microondas hasta la estructura a gran escala de galaxias y cúmulos. Sin embargo, estas observaciones no nos permiten concluir si el agente responsable de la aceleración actual es una constante o bien evoluciona ligeramente con el tiempo. En un futuro próximo, con cartografiados más precisos, a bordo de satélites como Euclid de la ESA, podremos explorar las posibles desviaciones respecto de una constante cosmológica y determinar si existe algún campo cósmico cuya energía actual acelere el universo. Tales campos no son impensables, de hecho aparecen a menudo en la física de partículas, como por ejemplo el campo de Higgs, cuya densidad de energía es en principio enorme y es necesario un

mecanismo que la restrinja o cancele para que hoy en día tenga un valor compatible con las observaciones. Otro campo parecido es el del inflatón, cuya energía es responsable de la aceleración primordial del universo, que actuó durante una fracción infinitesimal de segundo, poco después del origen del universo, y que dio lugar a toda la materia y energía que observamos hoy en día, en un universo homogéneo y euclídeo. Dichos campos podrían estar relacionados con el misterio de la energía oscura, aunque por el momento es prematuro asignarle un valor a dicho campo. Esperemos que la exploración del universo con satélites como Euclid nos desvele el misterio.

Notas:
[1] Ver capítulo 53
[2] Ver capítulo 59

Bibliografía:
"La materia oscura", Alberto Casas, RBA (2015).
"El corazón de las tinieblas: materia y energía oscuras", Simon Mitton y Jeremiah Ostriker, Ed. Pasado y Presente (2014).

Juan García-Bellido Capdevila
Doctor Física Teórica
Catedrático, Universidad Autónoma de Madrid

Capítulo 53
A la caza de la materia oscura: viendo lo invisible
(Por Miguel Peiró)

Como hemos visto en el capítulo anterior nuestro Universo posee una complejidad que va mucho más allá de aquella presente en astros tales como estrellas o planetas, o de las mismas galaxias que los albergan. Una cantidad nada desdeñable de materia oscura lo impregna todo, está presente en los vastos espacios que existen entre cúmulos de galaxias, pero también está cruzando en este mismo momento su habitación, y a usted mismo, a una velocidad de unos cientos de kilómetros por segundo.

La certeza acerca de la existencia de este nuevo tipo de materia exótica es incontestable. A día de hoy se tienen numerosas observaciones que apuntan en una misma dirección: aproximadamente el 85% de la materia total del Universo se encuentra en forma de materia oscura, mientras que el 15% restante estaría compuesto de materia ordinaria (bariónica) o lo que es lo mismo, la materia que hasta ahora conocíamos, aquella compuesta de protones, neutrones y electrones. Sí, todo lo que le rodea, todo lo que se puede observar en el cielo en una noche estrellada, todas las galaxias, planetas y estrellas, no son más que una parte pequeña de todo lo que *ahí* fuera nos aguarda.

Sin embargo, las numerosas observaciones que prueban la existencia de la materia oscura solo prueban que ésta interacciona gravitacionalmente con la materia ordinaria. Es decir, de la misma forma que el Sol y la Tierra sienten una atracción gravitatoria mutua, la materia oscura siente esa atracción hacia la materia ordinaria, y viceversa. Pero la gravedad es solo una de las cuatro fuerzas fundamentales que conocemos en la Naturaleza. En lo relativo a las otras tres, el electromagnetismo, la fuerza nuclear fuerte, y la fuerza nuclear débil, no se tiene constancia a día de hoy de que la materia oscura pueda interaccionar con la materia ordinaria a través de estas fuerzas. Es más, muy posiblemente la materia oscura no posea las propiedades necesarias para que pueda interactuar ni electromagnéticamente ni a través de la fuerza nuclear fuerte con la materia bariónica. Este exótico nuevo tipo de materia fue denominada oscura, aunque un término más acertado hubiera sido transparente, ya que no emite ni absorbe luz, que es precisamente una consecuencia de su nula (o realmente pequeña) interacción electromagnética.

Pero, ¿cómo es posible entonces observar la materia oscura? Su observación directa o indirecta es fundamental a la hora de desentrañar uno de los mayores misterios de la física moderna: la naturaleza de su composición. Pues bien, actualmente se cree que sí sería posible observarla directa e indirectamente, e incluso hasta sería plausible crearla aquí en nuestro planeta. Y es que si hay algo que sabemos con certeza, y que ya hemos mencionado anteriormente, es que la cantidad de materia oscura es mucho mayor que la cantidad de materia ordinaria y que por lo tanto, debió de existir algún

mecanismo en los orígenes del Universo que crease tal cantidad de materia oscura. Esta es una pieza clave a la hora de entender este puzzle.

Cuando un físico se sienta en su mesa, delante de un papel, dispuesto a revolucionar el mundo de la física con una nueva teoría que dé cuenta de la materia oscura de nuestro Universo, lo primero en lo que ha de reparar es en cómo crear esa enorme cantidad de materia que observamos. A día de hoy conocemos muchos mecanismos que explicarían estas observaciones, pero lo más importante es que en la mayoría de ellos el punto diferencial es el grado de interacción que se le supone a la materia oscura con la materia ordinaria. Y es que aunque al lector le pueda parecer extraño, a través de las interacciones entre ambos tipos de materia se puede calcular muy precisamente cuanta materia oscura habría en el Universo que poco a poco se va creando en nuestro papel. ¡Qué curioso!, resulta que si la materia oscura interacciona débilmente, y recalcamos que solo débilmente, con la materia ordinaria (con la intensidad correspondiente a la fuerza nuclear débil) la cantidad de materia oscura en nuestro Universo se correspondería a aquella que se ha inferido observacionalmente, ese 85% que previamente mencionábamos. Esto parece más que una coincidencia, es lo que se vino a llamar el milagro de los WIMPs (*Weakly Interacting Massive Particles*), ya que este tipo de partículas hipotéticas seguirían estas pautas. Pero las consecuencias de esto no se limitan a la cantidad de materia oscura. Si esto realmente fuera así, y la materia oscura estuviera compuesta por WIMPs, significaría que existen otras formas de verla, y esto es demasiado grande e importante para que pase desapercibido.

53.1 Detección directa.

Podemos ahora formularnos la siguiente pregunta: si para que se formara tanta materia oscura en el Universo ésta ha de interactuar débilmente con la materia ordinaria, y teniendo en cuenta que la materia oscura atraviesa la Tierra en enormes cantidades, ¿es razonable tratar de detectar esa interacción aquí en nuestro planeta? La respuesta a esta pregunta es sí. Esto ha motivado un programa experimental mundial para la detección de la materia oscura sin precedentes.

Esta técnica para la detección de la materia oscura se basa en la posibilidad de que una partícula de materia oscura choque contra un núcleo atómico, contra alguno de los protones o neutrones que lo compone. Como bolas de billar, tras el choque, el núcleo experimentará un retroceso y posteriormente liberará energía en diferentes formas. Es esta energía la que se trata de medir con tanto ahínco, y que a día de hoy no se ha conseguido.

La probabilidad de que una partícula de materia oscura choque, o lo que es lo mismo, interaccione con un núcleo de un detector es tremendamente pequeña. Es por esta razón que los experimentos de detección directa están compuestos de ingentes cantidades de un material, como Germanio o Xenon, que se deja durante largos periodos de tiempo a la caza de una de estas interacciones. Estas dos variables, la cantidad de material y el tiempo de *escucha*, resultan en lo que se denomina exposición. Cuanto mayor sea la exposición de un experimento dado, mayor es la probabilidad de que una partícula de materia oscura choque contra uno de los núcleos del detector, lo que hace que este tipo de experimentos traten incesantemente de aumentar las cantidades del material que lo compone. Sin embargo, no todo es tan fácil como puede parecer. La lluvia de partículas que recibe la Tierra no está ni mucho menos solo compuesta de

materia oscura, existen muchos otros tipos de partículas que también nos bombardean, y concretamente en este tipo de experimentos son de especial importancia lo que se conocen como los rayos cósmicos.

Los rayos cósmicos son radiación de muy alta energía proveniente en su mayoría de procesos astrofísicos fuera del sistema solar, como supernovas o quasares. Cuando alcanzan la atmósfera terrestre chocan contra moléculas que la componen produciendo en este proceso una lluvia de numerosas partículas como piones, neutrinos o muones, conocidos como rayos cósmicos secundarios. En su camino, estas partículas pueden chocar contra uno de nuestros detectores de materia oscura produciendo una señal *similar* a la que produce la materia oscura. Para aislarse de este tipo de choques, los experimentos de materia oscura se sitúan a grandes profundidades bajo tierra, usando así el propio suelo como escudo ante este tipo de intrusos. Antiguas minas o montañas son el lugar habitual en el que podemos encontrar grandes experimentos de detección directa de materia oscura como LUX o SuperCDMS.

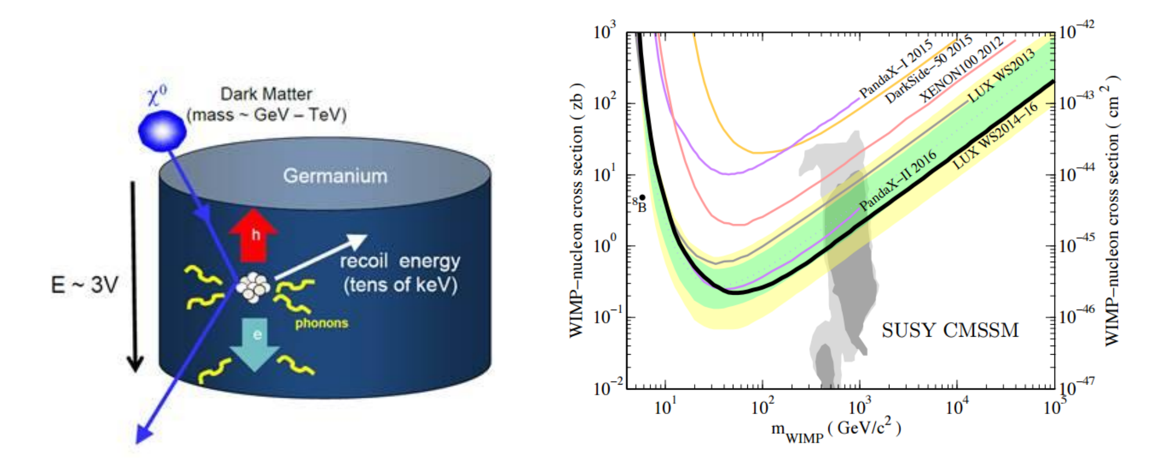


Figura 1: *A la izquierda, representación gráfica de una partícula de materia oscura chocando contra un núcleo de germanio, como puede ser en el experimento SuperCDMS. A la derecha, compendio de los últimos resultados obtenidos en el campo de la detección directa. En el eje y, la probabilidad de choque entre materia oscura y protón, en el eje x, masa de la materia oscura. Cualquier modelo cuya predicción de probabilidad de choque esté por encima de la línea negra quedaría descartado experimentalmente por detección directa (imagen obtenida de arxiv.org/abs/1608.07648).*

Como ya hemos mencionado anteriormente, hasta ahora no se han encontrado señales en detección directa que con toda seguridad hayan sido producidas por materia oscura. Lejos de producir un declive en el interés de la comunidad científica, esto ha servido para seguir aunando fuerzas en esta desconcertante caza. No obstante, la nula detección de materia oscura hasta ahora ha servido para entender que muchos modelos que se tenían no podrían ahora explicar este problema. Es decir, si tenemos un modelo que predice un determinado grado de interacción entre la materia oscura y los núcleos atómicos tal que la materia oscura debería haberse visto ya en un determinado experimento de detección directa, este modelo quedará automáticamente descartado, ya que como decíamos, aún no se ha observado nada.

53.2 Detección indirecta.

La detección directa, que acabamos de ver, no es la única manera que conocemos para probar las interacciones no gravitacionales de la materia oscura con la materia ordinaria. La detección indirecta de materia oscura es una técnica basada en las mismas hipótesis que anteriormente describíamos, pero que a diferencia de la detección directa, trata ver esas interacciones en lo que se conocen como halos de materia oscura. Se llama halo de materia oscura a toda esta materia que envuelve al disco de materia visible en las galaxias, como la nuestra, la Vía Láctea, y que se extiende mucho más allá de los límites visibles de ésta. Es decir, es toda la materia oscura que baña las galaxias, desde los centros de éstas hasta zonas muy externas que *a priori* no eran siquiera consideradas como partes de la propia galaxia. Tanta es la cantidad de materia oscura que existe en los halos, que la mayor parte de la masa total de una galaxia proviene de esta componente, e incluso se sabe que existen galaxias en las que la cantidad de materia oscura supera por muchísimo la cantidad de materia ordinaria, como es el caso de las galaxias enanas esferoidales (dSph, del inglés *dwarf spheroidal galaxies*). En galaxias como la nuestra, la densidad de materia oscura del halo crece de forma drástica a medida que nos acercamos al centro de esta, existiendo en esa región enormes densidades de materia oscura rotando alrededor del agujero negro supermasivo que allí se encuentra.

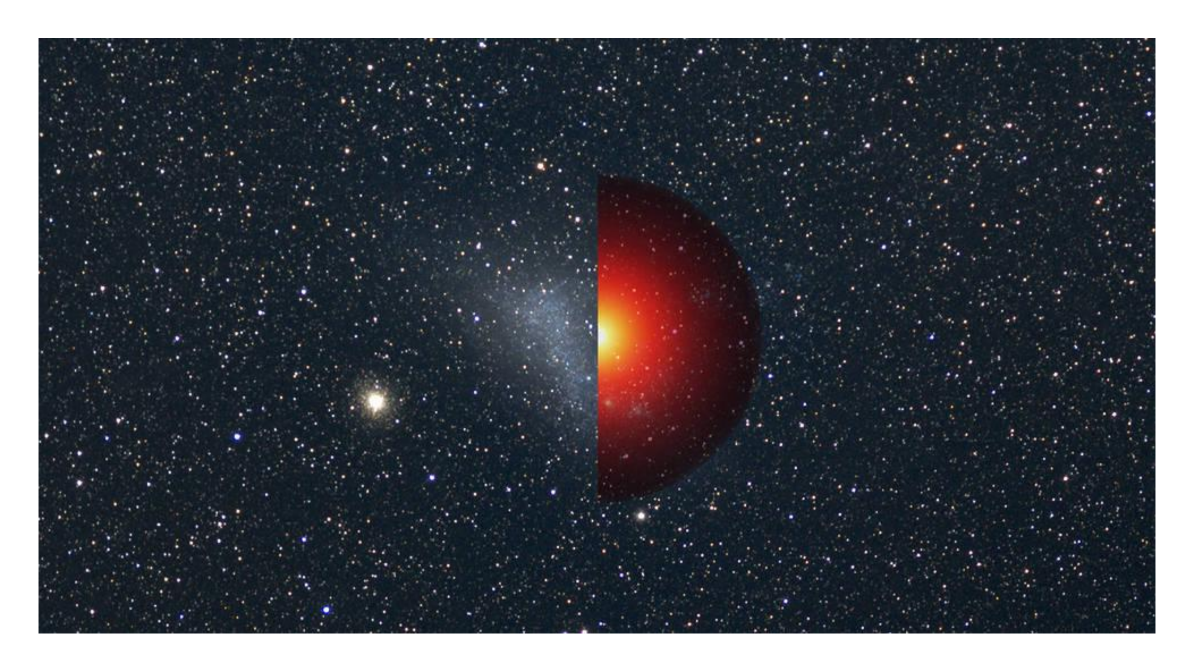


Figura 2: *Nube pequeña de Magallanes. La fotografía superimpone una imagen del visible (del centro hacia la izquierda) con un modelo de la distribución de materia oscura que ésta ha de contener (del centro a la derecha). La intensidad de colores refleja la densidad de materia oscura, mucho mayor en la zona central (imagen de http://www.nasa.gov).*

Por mágico que pueda parecer, hoy en día sabemos que, a nivel microscópico, el mismo proceso que genera el choque entre la materia oscura y la materia ordinaria (aquél que nos servía como motivación de los experimentos de detección directa) sería responsable de lo que se conoce como la *aniquilación* de la materia oscura. Cuando dos partículas de materia oscura se encuentran, existe una cierta probabilidad distinta de cero de que éstas se fundan dando lugar a un par partícula-antipartícula de materia

ordinaria, como por ejemplo, un par electrón-positrón o un quark-antiquark. Pues bien, estas partículas son las que se tratan de detectar mediante la técnica de la detección indirecta. En la mayoría de los casos, las partículas creadas no son estables, y acaban por decaer en otras partículas que si lo son, y por increíble que nos resulte, entre éstas hay fotones, partículas de luz. Entonces al fin y al cabo, la materia oscura sí genera luz, mediante su aniquilación. Pero la probabilidad de que esto suceda es relativamente pequeña y esto hace que, desgraciadamente, la materia oscura no sea una fuente lumínica que pueda ser comparada con la mayoría de procesos astrofísicos que ocurren en una galaxia, lo que dificulta mucho la tarea de su detección mediante esta técnica. Sin embargo, buscar estas aniquilaciones en zonas donde la densidad de materia oscura es grande, como el centro de la Vía Láctea, o dónde prácticamente toda la luz generada pueda venir de la materia oscura, como en las dSph, ayuda a maximizar las posibilidades de un descubrimiento.

A pesar de esta dificultad existen numerosos experimentos que han aceptado este magnífico reto. Satélites como Fermi-LAT están tratando de cazar estos fotones provenientes de las aniquilaciones de materia oscura, y otros como AMS tratan de detectar la antimateria producida (positrones por ejemplo). Los detectores de neutrinos también están al tanto de la posibilidad de detectar una fuente de neutrinos provenientes de estas aniquilaciones (los neutrinos, como otras partículas que ya hemos mencionado, se pueden producir en el choque de dos partículas de materia oscura), y redes de telescopios como CTA buscarán los rayos cósmicos producidos en la atmósfera cuando partículas generadas por las aniquilaciones la crucen. En definitiva, el campo de la detección indirecta es, actualmente, un campo de mucha actividad. Desgraciadamente, no existen aún indicios de una detección clara de la materia oscura en este campo. Pero como en el caso de la detección directa, esto también nos da información muy valiosa acerca de la naturaleza de esta exótica forma de materia.

53.3 LHC.

Quizá llegados a este punto alguien haya pensado lo siguiente: si al encontrarse dos partículas de materia oscura generan, por ejemplo, un par quark-antiquark, ¿no puede ocurrir el proceso inverso?, es decir, que un par quark-antiquark dé lugar a un par de partículas de materia oscura. Esto sería algo así como crear materia oscura a partir de materia ordinaria, y la respuesta es un sí rotundo. Además, para que esto ocurra, no hace falta irse tan lejos como el centro de la galaxia, esto podría estar sucediendo muy cerca de nosotros en los aceleradores de partículas.

Los aceleradores de partículas son máquinas que mediante el uso de campos electromagnéticos aceleran partículas a velocidades cercanas a la de la luz. En el LHC (*Large Hadron Collider*), el acelerador más grande y potente jamás construido, se aceleran dos protones en direcciones opuestas para más tarde hacerlos colisionar. En el momento de la colisión se generan infinidad de partículas, pero lo más interesante en el contexto de las búsquedas de materia oscura es que se crean pares quark-antiquark que podrían dar lugar a la creación de materia oscura.

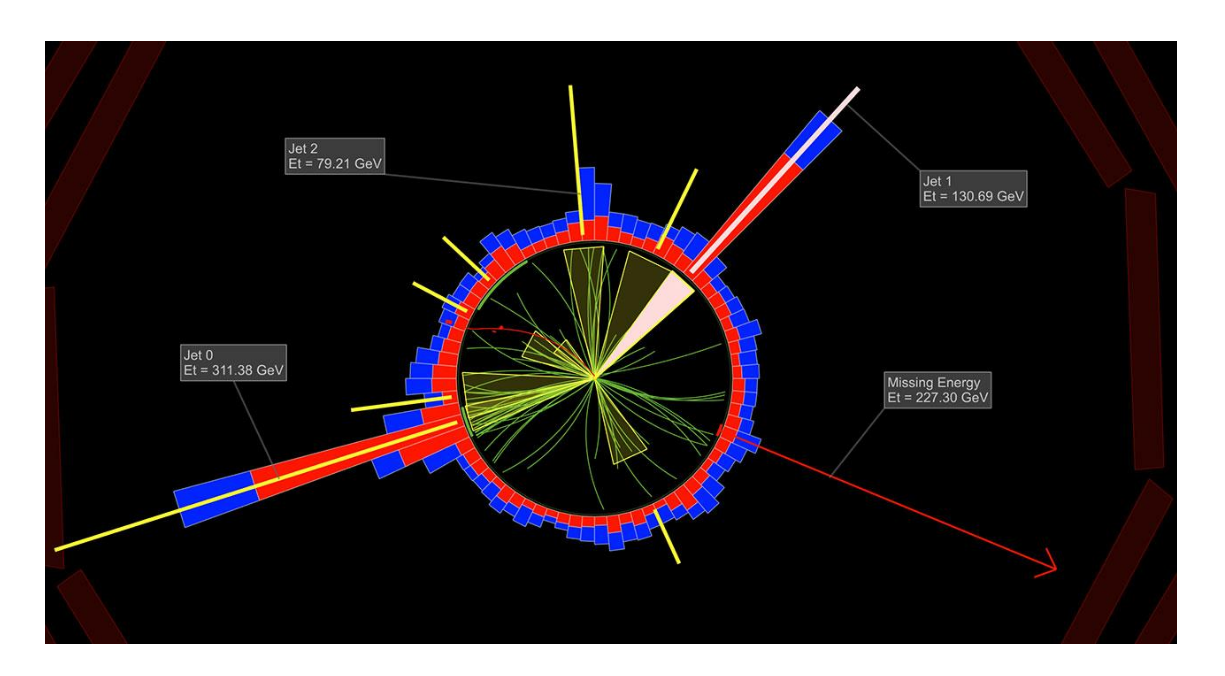


Figura 3: *Vista frontal de una colisión en el LHC. La creación de materia oscura está representada por la flecha roja de la derecha, y crea un balance negativo de la energía final (imagen de http://www.sdsc.edu/).*

En este caso el proceso de aniquilación ocurre en la materia ordinaria y el producto de ésta es materia oscura. Pero como ya sabemos, aunque la materia oscura interaccione con la materia ordinaria, la intensidad de esta interacción ha de ser muy pequeña, es decir, que no solo será poco probable crearla a través de aniquilaciones sino que además cuando esto ocurra, la materia oscura saldrá disparada sin que haya ninguna posibilidad de detectarla. Sin embargo, las leyes de la física están de nuestro lado. Sabemos que la energía se conserva, esto es, toda la energía que lleven los dos protones que chocan (una cantidad fácilmente calculable) ha de ser la misma que toda la energía que se exista tras la colisión. Esto significa que cuando se cree materia oscura y escape a nuestra detección (en este caso el proceso sería exactamente el mismo a la detección directa) lo que vamos a ver es un balance negativo entre la energía inicial y la energía final. Es decir, la materia oscura se detecta por la energía que se lleva.

Tampoco a día de hoy se ha detectado la presencia de materia oscura en el LHC. Pero, de nuevo, esto ha servido para descartar algunas posibilidades y diseñar nuevas estrategias para la detección de materia oscura en aceleradores de partículas.

53.4 Conclusiones.

La naturaleza de la materia oscura ha supuesto uno de los mayores retos de la física moderna. A pesar de estar presente en enormes cantidades en el Universo, y de atravesar la Tierra incesantemente, aún a día de hoy y tras cuatro décadas de investigación, no sabemos de qué está compuesta. Como en un juego casi maquiavélico la naturaleza nos enseña una pequeña parte de su vasta y maravillosa complejidad, pero se guarda bajo llave el resto.

Hemos visto que la tarea de detectar la materia oscura es realmente complicada y está llena de interrogantes. Aún con todo esto, el ser humano ha sido capaz de diseñar potenciales estrategias de detección que nos doten de ojos para ver el lado oscuro del Universo. No hemos sido capaces aún de dar con la tecla, pero el programa experimental en este sentido no hace más que crecer, y lo que es quizá más importante, la comunidad científica no para de aprender e investigar para dar respuesta a todos estos interrogantes.

Es posible que haya que esperar aún muchos años hasta que se detecte la materia oscura, o es posible que se detecte mañana mismo, quién sabe. Lo que es seguro es que su detección supondrá un hito sin precedentes en la historia de la Ciencia y de la Humanidad.

Miguel Peiró
Doctor en Física Teórica
Doctor Contratado, Universidad Autónoma de Madrid

Capítulo 54
¿Existe realmente la materia oscura?

(Por Xavier Hernández y Christine Allen)

El estudio científico de la gravedad comenzó a finales del siglo XVI con los trabajos de Galileo Galilei, quien dedicó largas horas a observar detenidamente el movimiento de objetos bajo el efecto de esta fuerza. Una de sus observaciones más relevantes, fue la que estableció que "un kilo de plumas cae igual de rápido que un kilo de plomo" Para decirlo en otras palabras, se dio cuenta, dejando caer objetos de distintos materiales por pequeñas rampas de madera y soltándolos desde la Torre de Pisa, que el material del que un objeto está hecho no afecta su movimiento bajo los efectos de la gravedad. Galileo identificó correctamente que es la fricción con el aire lo que ocasiona que una pluma caiga más lentamente que una piedra sobre la superficie de la Tierra. Posteriormente se verificó esta observación mediante experimentos con tubos al vacío.

Hacia finales del siglo XVII, Isaac Newton sistematizó y extendió los trabajos de Galileo, para llegar a una descripción bastante exacta de cómo opera la gravedad dentro del Sistema Solar. Newton propuso una "Ley Universal" según la cual la fuerza entre cualquier par de objetos es proporcional al producto de las masas de éstos, e inversamente proporcional al cuadrado de la distancia que los separa. Es interesante ver que para Newton, el sentido de la universalidad de esta ley se debía a que era válida para todos los objetos del universo, pero no necesariamente a que mantuviera una forma única en todos los ambientes. De hecho, en ausencia de una deducción lógica, exploró varias formas para esta ley, y se quedó con la dependencia conocida, dado que era la que explicaba bien los movimientos medidos en la Tierra y en todo el Sistema Solar entonces conocido. De esta manera, la gravitación de Newton es un ajuste matemático a una serie de observaciones empíricas. Uno de los más notables logros de esta teoría fue la predicción del regreso del cometa que hoy llamamos de Halley. Antes, la aparición de los cometas se consideraba una anomalía a las regularidades del Sistema Solar. Lo variado de la apariencia de los cometas en pasos sucesivos cerca del Sol, su gran número y sus órbitas de períodos largos no permitieron identificar su regularidad, hasta que ésta se dedujo como una consecuencia de la gravedad de Newton.

Durante el siglo XVIII se exploraron los detalles de la teoría de Newton, notablemente por Laplace y Lagrange, y se llegó a construir un "Universo de relojería", donde el termino "Universo" se usaba para referirse al Sistema Solar. Todos los cuerpos del Sistema Solar se movían en trayectorias precisas según las expectativas de la teoría de Newton. Hacia mediados del siglo XIX se detectó una interesante anomalía a la regla: el planeta Urano parecía a veces retrasarse ligeramente en su órbita, y luego adelantarse. Suponiendo la validez rigurosa de la teoría de Newton, Adams y LeVerrier, independientemente, llegaron a la conclusión de que esta anomalía se debía a la presencia de un planeta extra, que se movía en una órbita externa a Urano. La teoría de

Newton predecía entonces la posición de este nuevo objeto, que efectivamente se encontró justo donde tenía que estar: se había descubierto un nuevo planeta, Neptuno, en 1846.

Posteriormente, hacia finales del siglo XIX, se puso de manifiesto otra anomalía gravitacional. Esta vez, el planeta más interno conocido se comportaba de manera extraña. La órbita de Mercurio no se cerraba, sino que, lentamente, el perihelio de su órbita avanzaba, órbita tras órbita. Parte de este efecto se podía entender mediante la gravedad newtoniana, si se incluía cierto achatamiento del Sol y las perturbaciones de los planetas conocidos, pero no era suficiente para explicar el avance del perihelio observado. Se propuso entonces la existencia de un nuevo planeta interno a Mercurio, Vulcano. Aunque su observación fue reportada varias veces, este planeta de hecho no existe. No fue sino hasta principios del siglo XX que se entendió el corrimiento del perihelio de Mercurio, no como debido a la presencia de una masa no observada, sino como una indicación de que la teoría Newtoniana no era una explicación perfecta de la gravedad. Fue con la teoría de la relatividad general de Einstein que se explicó la anomalía de la órbita de Mercurio, en términos de un sistema que se encuentra fuera de la región de validez de la teoría dominante, la gravedad de Newton. Así, se hizo evidente la necesidad de extender el marco teórico de la física de esa época. En la relatividad general, los movimientos no rectilíneos de objetos bajo el efecto de la gravedad no se deben a una fuerza de atracción, sino a la curvatura que en el espacio ocasiona (cambios en la medida de intervalos espaciales y temporales que varían con la posición). Por ende, según la observación de Galileo, todos los cuerpos se mueven de la misma manera sin importar su composición, ya que se están moviendo en el mismo espacio curvado.

De esta manera, vemos que anomalías gravitacionales, a veces, han llevado a el descubrimiento de materia “oscura" previamente no identificada (como en el caso de Neptuno), y a veces, han revelado que las teorías existentes son solo aproximaciones válidas en un intervalo limitado de parámetros físicos (como en el caso del corrimiento del perihelio de Mercurio).

Durante el siglo XX, el intervalo de distancias a las que se estudian los sistemas astronómicos creció enormemente. Desde mediados del siglo XIX se sabía que el Sistema Solar medía solo una fracción ínfima de lo que mide nuestra galaxia, y en los años de 1920 se estableció que las "nebulosas espirales" eran de hecho "universos islas", para usar los términos de la época, esto es, galaxias externas parecidas a la nuestra, pero a distancias decenas, centenares y miles de veces mayores que el tamaño de nuestra Vía Láctea. Para entonces, se tenían ya dos formas independientes de estimar la masa de un sistema astronómico: midiendo cuánta luz emite, y estudiando su dinámica. Dado que el Sol es una estrella más o menos promedio, si vemos que un pedazo de galaxia emite 1000 veces más luz que el Sol, en primera aproximación pesará 1000 veces más que éste. Por otro lado, asumiendo la validez de la relatividad general (o equivalentemente para velocidades mucho menores que la de la luz, la teoría de Newton), si observamos los tamaños y las velocidades típicas de un sistema astronómico, podemos deducir cuánta masa es necesaria para mantener al sistema unido y estable. Por lo general, a escalas sub-galácticas, (por ejemplo, para los cúmulos de estrellas) ambas determinaciones coinciden dentro de sus errores, y todo cuadra.

Sin embargo, al considerar sistemas cada vez más grandes, la situación cambia. Ya en los años 1930, Fritz Zwicky obtuvo un resultado sorprendente: la dinámica de

enormes cúmulos de galaxias implicaba, según la relatividad general, que estos contenían unas 20 veces más masa de la que se veía en las galaxias que pertenecían al cúmulo. Zwicky propuso la hipótesis de la materia oscura, que para él representaba materia normal que no brillaba, y que podía ser gas muy caliente, polvo, planetas, etc. De hecho, durante décadas, la búsqueda de la materia oscura fue una de las principales motivaciones para ampliar el espectro electromagnético estudiado por los astrónomos. Si la materia no brillaba en el óptico, el angosto intervalo de radiación al que Zwicky tenia acceso, seguramente se detectaría en radiofrecuencias, en las microondas, en el infrarrojo, en los rayos X, o en algún otro intervalo espectral.

Hacia los años 1970, los estudios dinámicos en galaxias espirales hechos por Vera Rubin y los radioastrónomos holandeses revelaron que estas galaxias tenían más o menos el mismo faltante de masa que el inferido por Zwicky para los cúmulos de galaxias, bajo la suposición de que la anomalía gravitacional detectada se debía a un faltante de masa, una hipótesis “ad hoc” para forzar la concordancia entre las observaciones y la mecánica de Newton. Hay que recordar que Mercurio se encuentra a una distancia del Sol apenas diez veces menor que la Tierra, mientras que la dinámica de galaxias y cúmulos de galaxias ocurre a escalas miles de millones de veces mayores que el tamaño total del Sistema Solar.

Durante los años 1980 y 90, al estudiar la dinámica de la expansión del Universo, se llegó a la conclusión de que a las escalas más grandes observadas también se tenía una fuerte disparidad entre la cantidad de materia observada y la requerida por la teoría de la gravedad de Einstein para explicar las observaciones. Se cimentó la idea de que este faltante de masa (de masa capaz de emitir luz) era real, e incluso, se empezó a presentar como un hecho científicamente comprobado. ¡Solo faltaba detectar este componente directamente!

Ya para esa época, había estudios que abarcaban todo el espectro electromagnético (muchas de ellos desde satélites) y que habían dejado en claro que la "materia oscura" no brillaba en ninguna longitud de onda. Durante algún tiempo, los neutrinos (partículas subatómicas con una bajísima probabilidad de interacción con los átomos) fueron el candidato favorito para la "materia oscura". Para empezar, ¡tienen la gran virtud de que existen! Poco a poco, las mediciones directas de la masa de los neutrinos dejaron en claro que éstos no podían ser más que una fracción mínima de la materia oscura, y se pasó entonces a considerar partículas hipotéticas predichas por modelos de supersimetría en física nuclear. La detección de estas partículas se ha anunciado espectacularmente de tanto en tanto, pero estudios subsecuentes han acabado siempre por descartar los múltiples anuncios. Estos resultados negativos ya no han recibido la correspondiente publicidad.

Durante los últimos 30 años se han invertido muchísimos millones y muchísimos esfuerzos en buscar directamente a estas partículas, miles de las cuales tendrían que estar cruzando cada centímetro cuadrado del área de la Tierra cada segundo, y que constituirían así cerca del 90% de la masa en el Universo. Gradualmente, ha sido forzoso descartar a los candidatos más plausibles, lo que ha llevado a propuestas cada vez más exóticas, en especial ahora, después de que en el verano de 2016 experimentos en el LHC en CERN descartaron cualquier partícula viable dentro de los modelos de supersimetría estándar.

De esta manera, estamos ante una situación en la que el agente causal determinante de nuestra teoría de gravitación se postula como un elemento que nunca nadie ha visto, situación que recuerda el caso del éter de finales del siglo XIX. Lo anterior ha generado un interés creciente en las teorías de gravitación modificada, en las cuales se propone un estudio más empírico del problema, tratando de inferir el carácter de la gravedad a escalas astronómicas, muchos órdenes de magnitud mayores que las escalas accesibles a experimentos directos, y partiendo de la condición de explicar la dinámica observada usando solo la materia detectada.

No existe todavía una teoría de gravedad modificada única o final, refinada y determinada, al nivel de la gravedad de Einstein. Pero ya contamos con una serie de primeras aproximaciones, relaciones de escala y reglas empíricas que están orientando este desarrollo. Mediante ellas, se ha explorado un número cada vez mayor de sistemas dinámicos cuya explicación tradicional requiere de la presencia de materia oscura. Se encuentra de manera consistente, que las predicciones de la gravedad modificada (en su forma empírica actual) se confirman. Siempre que un sistema dinámico se encuentre en un régimen de aceleración unas mil millones de veces menor que la aceleración gravitacional sobre la Tierra (y solo en esos casos), se presentan las anomalías gravitacionales que según la gravedad estándar se identifican como evidencia indirecta de materia oscura.

La gravedad modificada ha probado, en los sistemas a los que ha sido aplicada, que la existencia de materia oscura es innecesaria. Si bien aún se encuentran en su infancia, las teorías de gravedad modificada permiten ya vislumbrar la posibilidad de que la búsqueda de las exóticas partículas responsables de cerca de 90% de la masa del universo sufra la misma suerte que la búsqueda del éter en el siglo XIX.

Christine Allen Armiño **Xavier Hernández Döring**
Doctores en Astrofísica
Investigadores
Instituto de Astronomía, Universidad Nacional Autónoma de México

Capítulo 55
La fisión nuclear ¿sigue avanzando o tiene fecha de caducidad?
(Por Ignacio Durán Escribano)

La respuesta rápida a esta cuestión no puede ser otra que la de constatar su avance y pronosticarle una centenaria trayectoria futura. Pero es cierto que se trata de un tema controvertido y merece pues un análisis más detallado.

No es este el lugar de discutir acerca de los motivos que puedan tener – que tienen – los detractores de la energía nuclear y merecería un artículo separado discutir acerca del riesgo real y de la percepción que de ese riesgo tiene la población de cada país. Voy a asumir que la controversia existe, pero partiendo del hecho constatable de que a día de hoy hay 448 reactores operativos, 58 en construcción, 167 programados y 345 en fase de estudio, según los datos del *Nuclear Fuel Report* de la World Nuclear Association. De estos números hay que deducir los 40 reactores japoneses que permanecen parados desde el accidente de Fukushima. Desde 1996, 70 reactores han sido definitivamente cerrados y solo 76 nuevos han sido puestos en funcionamiento, pero la energía eléctrica generada se ha prácticamente duplicado, debido a que muchos de los reactores en operación se han renovado, alargando su vida útil y aumentando su eficiencia energética, además de que la mayor parte de los retirados tenían una potencia nominal entorno a los 600 MWe mientras que los nuevamente construidos tienen una potencia media cercana a los 1000 MWe y ciclos de funcionamiento más extendidos. Es decir, una visión objetiva del estado actual es la de un sector en crecimiento, con la previsión de que para 2035 se cierren 132 reactores mientras que se estima que se pondrán en funcionamiento 290 nuevos.

¿Qué hay de nuevo en la tecnología nuclear de fisión y como se adapta a la evolución social? No veo yo que se pueda hablar de grandes cambios en los diseños básicos, si bien hay un consorcio, establecido al más alto nivel internacional, para el estudio y desarrollo de nuevas tecnologías, que se engloba bajo el título de GEN IV (cuarta generación). En el momento actual la actividad principal se centra, sin embargo, en el desarrollo de la denominada tercera generación, a la que se le añaden términos para reforzar el concepto de avance continuo en los resultados obtenidos. Así se habla de GEN III+ o se antepone el adjetivo de *Advanced* al mencionar la tecnología empleada y se observa la creciente atención, tanto por la reducción de costes como por el aumento de la seguridad de las instalaciones. Estos reactores tienen mejor eficiencia térmica que los construidos hace veinte años y alargan los tiempos de funcionamiento entre paradas programadas, con el consiguiente aumento de la energía generada año tras año. Hay que decir aquí que estos reactores de nueva generación nacen con tiempos de vida útil de sesenta años, no estando excluido que puedan luego reformarse para duplicar ese valor.

Tres han sido los hitos que han marcado la orientación de los nuevos diseños, el trágico accidente de Chernobil, los acuerdos de Kioto sobre el cambio climático y el accidente de Fukushima tras sufrir el impacto de un extraordinario tsunami. Los dos graves accidentes tendrían, supuestamente, un impacto negativo en el desarrollo de la energía nuclear. Pero no ha sido así y tras cada uno de estos sucesos se observa un parón en la construcción de nuevos reactores, seguido unos años después de un repunte, sobre la base de diseños orientados, tanto hacia la mayor seguridad como hacia el mejor rendimiento energético. Los acuerdos derivados de la toma de conciencia del calentamiento global conducen a un replanteamiento del mercado de la energía y la racionalización que de él se deriva ha dejado clara la necesidad de definir lo que viene llamándose el *mix* energético, en el que la energía nuclear tendría asignado un porcentaje en torno al 20%, según los distintos escenarios contemplados. El cambio fundamental a lo largo de estos años ha sido el pasar de un planteamiento local y puramente económico a otro basado en la seguridad del suministro eléctrico, con la mínima producción de CO_2 y el mínimo riesgo para la población, a escala mundial.

Entre tanto, la incorporación al primer nivel de desarrollo de las grandes potencias orientales aporta, no solo un incremento en la demanda de energía sino también una oferta de su capacidad constructiva en el sector nuclear. Esto ha provocado la formación de grandes consorcios industriales internacionales, liderados por Westinghouse Electric (que pasa, de estar controlada por capital británico a estarlo por Toshiba y participada con otras compañías orientales) y GEH (General Electric + Hitachi), junto con las grandes empresas nacionalizadas de Francia, Rusia o Korea. Este movimiento industrial y financiero ha propiciado que el diseño de los nuevos reactores se haga de forma modular, es decir, que su construcción sea hecha por partes, en distintas empresas e incluso en distintos países y, de esta forma, un número creciente de empresas de muy distintos lugares pueden licitar a la hora de formarse cada uno de los consorcios adaptado a cada planta concreta. El resultado de esta dinámica industrial es una reducción de los costes y de los tiempos de construcción y, al mismo tiempo, desaparece la total dependencia que antes existía de las tres grandes potencias del sector, USA, Rusia y Francia.

El planteamiento económico de la energía nuclear condujo al diseño de reactores cada vez más grandes, con una repercusión directa sobre el elevado coste financiero que conllevan los macroproyectos. Esto entraña compromisos políticos, no siempre fáciles de asumir en una sociedad cuya percepción del riesgo asociado a las centrales nucleares es muy desfavorable. La nueva conciencia de la necesidad de contribuir al *mix* energético y el incremento de la seguridad de los nuevos diseños, ha hecho cambiar los planteamientos, lo que se manifiesta más claramente en los países en vías de un fuerte desarrollo socio-económico, obligados a tomar decisiones sin excesivas concesiones a la presión mediática. Vemos así que, frente a la tímida actividad en Europa (4) y USA (4), la mayor parte de los reactores en construcción están en China (20), Rusia (7), India (5), Pakistan (2), los Emiratos Àrabes (4) y Korea (3).

A nivel de la seguridad, los accidentes de Chernobil y Fukushima, pero sobre todo este último, han conducido a una revisión de los procedimientos, con un papel creciente de la Agencia Internacional de Viena (la IAEA). Los distintos agentes involucrados en la construcción de los nuevos reactores tienen ahora especificaciones y controles más rigurosos, no solo por las agencias de seguridad nacionales sino también por la propia IAEA, lo que ha llevado a una mayor racionalización del proceso. Todos los nuevos diseños incluyen la redundancia de los sistemas críticos, la extrema

protección frente a catástrofes naturales y ataques terroristas y, por encima, lo que se denomina *seguridad pasiva*. Es decir, la tecnología empleada tiene que ser tal, que el eventual fallo de los sistemas de refrigeración no conduzca a la fusión del núcleo del reactor, permitiendo plazos de intervención razonables y suficientes, como para proceder a una parada segura. Todos los diseños actuales de GEN III que llevan la etiqueta *Advanced* cumplen con estos requisitos.

Dentro de las críticas que suelen hacerse a la tecnología nuclear están las dudas acerca de las reservas de uranio y torio, primero por su eventual escasez y segundo por la dependencia que se crea de los países consumidores respecto a los países productores. El análisis de la seguridad del suministro del combustible juega un papel muy importante a la hora de asumir el esfuerzo financiero que conlleva la construcción de una nueva planta de energía nuclear. Conduce también a un estado de ansiedad social, al pensar en el efecto negativo que sobre la economía de un país puedan tener los incrementos de los precios del combustible en el mercado internacional. Esta duda llegó a ser muy razonable y las respuestas no fueron muy favorables durante las pasadas décadas, al estimarse que las reservas de uranio y torio solo podrían garantizar precios estables durante apenas medio siglo. A corto plazo, esta dinámica cambió con el fin de la guerra fría y la introducción en el mercado de las reservas de plutonio y uranio procedentes de la actividad militar, por parte, fundamentalmente, de Rusia y USA. Por otra parte, a medio y largo plazo, resultará rentable el reprocesado del combustible usado, lo que conduce a pensar que las reservas disponibles permitan desbordar, al menos, el horizonte de varios siglos. Para entender esta afirmación hay que considerar que en la generalidad de los reactores funcionando actualmente, se aprovecha menos del 1% del potencial combustible, debido al aumento progresivo de la proporción de isótopos no fisibles, que capturan los neutrones y reducen la eficiencia de la reacción en cadena. En el ciclo actualmente normal del combustible, el aumento del stock de los actínidos de vida media larga (miles de años) plantea un grave problema, puesto que los residuos así generados tienen que ser almacenados en silos seguros y a la larga determinan el límite de la utilización de este tipo de reactores nucleares. No obstante, en los llamados reactores rápidos los neutrones no son moderados y alcanzan energía suficiente como para hacer fisionar todos los actínidos, produciéndose así la transmutación de estos isótopos que no son fisibles por los neutrones lentos. Estos reactores se construyen hoy en día de forma que son económicamente rentables, lo que permite afirmar que el coste total del combustible de fisión nuclear, incluyendo los reprocesamientos y la quema de los actínidos, seguirá siendo competitivo con las fuentes de energía renovables durante mucho tiempo.

En un resumen de lo aquí expuesto, se podría decir que la presión social contra los reactores nucleares de fisión hace que su uso se desplace hacia los países con fuerte crecimiento de la demanda energética, principalmente los orientales; que el alto grado de seguridad y la fiabilidad de los diseños hace previsible su funcionamiento por más de sesenta años; que los costes y tiempos de construcción disminuyen y su mayor eficiencia y vida útil reducen el coste de mantenimiento; que el reprocesamiento del combustible hace que se aprovechen unas veinte veces mejor los recursos actuales y que, finalmente, la construcción de reactores rápidos permite quemar los residuos de larga vida media, de forma rentable. Todo ello hace pensar que estamos hablando de una tecnología que va a convivir con las renovables durante, al menos, lo que queda de siglo.

En lo que respecta al *mix* energético ideal, no podría yo decir, a partir de lo que he leído, que el peso de la energía nuclear sea necesariamente el 20%. Hay un argumento en favor de que nunca dejase de ser mucho menor de ese valor, basado en la intermitencia de la mayor parte de las energías renovables y en el encarecimiento de la red eléctrica si hubiese que recurrir a un almacenamiento de la electricidad en grandes proporciones. Hay también consideraciones de orden económico, pero están basadas en diferentes escenarios del desarrollo social a nivel global y esto hace que las horquillas de las previsiones no conduzcan a estimaciones demasiado precisas. Si reducimos pues las previsiones a tiempos inferiores al medio siglo, vemos que la creciente implantación de las energías renovables – aun proyectando una tasa de crecimiento mayor que la actual – apenas cubrirá el aumento de la demanda de electrificación a nivel global. Es decir, si los componentes *limpios* (los que no producen CO_2) del *mix* energético no crecen a una tasa mucho mayor que la demanda, la deseada reducción del efecto invernadero no se producirá en estos cincuenta años, porque se mantendrán prácticamente constantes las componentes *sucias* (las basadas en los combustibles orgánicos).

Permítaseme terminar con una consideración acerca de la necesidad de fomentar el desarrollo en las regiones más desfavorecidas. Dentro de la tecnología actual están desarrollándose proyectos de reactores englobados bajos las siglas SMR (*Small Modular Reactors*) basados en la experiencia acumulada en los submarinos nucleares, cuya seguridad y fiabilidad está fuera de toda duda. Estos SMR se construyen en origen, destinándose a su utilización "llave en mano" en lugares remotos, sin necesidad de intervención a nivel local. Están previstos para su instalación allá donde se requiera un aporte de energía, sea electrificación o producción de agua desalinizada, de forma rápida, adaptando el número de módulos a las necesidades coyunturales, o en lugares de difícil acceso, como islas o desiertos. Pues bien, este tipo de reactores está llamado a ser una herramienta fundamental en la ayuda al desarrollo del tercer mundo.

Ignacio Durán Escribano
Doctor en Física
Catedrático, Universidad de Santiago de Compostela

Capítulo 56
¿Se llegará a obtener energía gracias a la Fusión Termonuclear controlada? ¿Es una cuestión tecnológica, económica o política?
(Por Julio Gutiérrez Muñoz)

Antes de comenzar a analizar estas preguntas, conviene dar unas pinceladas sobre qué es la Fusión Termonuclear y su nomenclatura habitual.

Las altas temperaturas necesarias para vencer la repulsión eléctrica de los núcleos a fusionar para obtener energía –millones de grados Celsius– hacen que los "combustibles" (normalmente Deuterio y Tritio) estén ionizados. Es decir, los átomos habrán perdido sus electrones y se encuentran en un estado denominado plasma, cuyas propiedades y comportamiento se conocen muy mal. Además, esas mismas temperaturas imposibilitan el confinamiento del plasma en recipientes, los cuales se volatilizarían al contacto con el gas ionizado. Actualmente, existen dos posibilidades, con ventaja sobre las demás barajadas hasta el momento, de controlar el proceso:

1) La **fusión por confinamiento magnético**, consistente en utilizar fuertes campos magnéticos para mantener el plasma en el interior de un recipiente, pero alejado de sus paredes.

2) La **fusión por confinamiento inercial**, también llamada **fusión por láser**, consistente en comprimir en el vacío la mezcla de núcleos a fusionar hasta densidades extremas, mediante el bombardeo de haces de luz procedentes de un láser de gran energía.

Ambos tipos de dispositivos, en su evolución desde los comienzos de la fusión nuclear, han adquirido tamaños descomunales como acreditan las figuras 1 y 2, y ese es el primer problema –quizás el más grave– que aleja de la realidad práctica la obtención de la energía prometida.

Ante la complejidad de la fusión, sea cual sea el tipo de confinamiento, ¿qué cabe decir? En consecuencia, los avances tecnológicos tienen, y seguramente tendrán por incontables años, mucho de acierto/error. Algunos de los problemas que presentan los diseños actuales tienen su origen, fundamentalmente, en la ingeniería y, en comparación con las dificultades conceptuales de la física subyacente, son extremadamente graves y difíciles de resolver.

Existe un problema adicional no menor. En el proceso se producen neutrones de alta energía imposibles de frenar antes de que atraviesen las paredes del dispositivo. Ello convierte en radiactivas las estructuras, por lo que se hace necesario encontrar materiales de baja activación frente al intenso bombardeo neutrónico, amén de otras partículas que escapan del confinamiento.

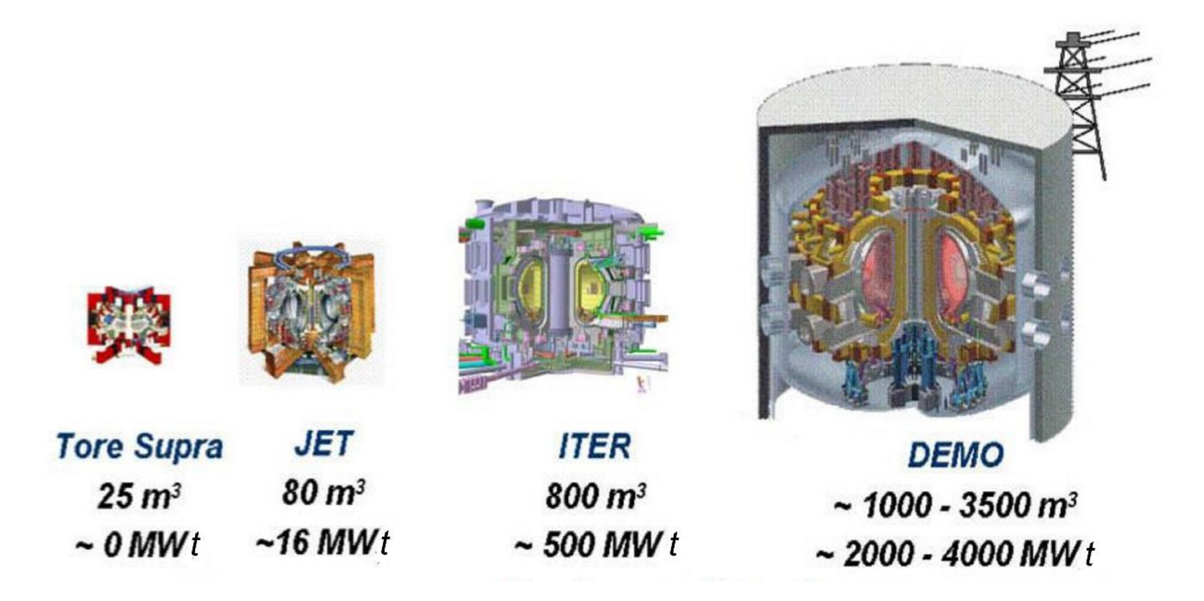


Figura 1: *Esquemas a escala que ilustran la evolución de los tamaños de los dispositivos de confinamiento magnético, los llamados "TOKAMAK" –acrónimo del ruso* ***то****роидальная* ***ка****мера с* ***ма****гнитными* ***к****атушками (****to****roidal'naya* ***ka****mera s* ***ma****gnitnymi* ***k****atushkami, en español cámara toroidal con bobinas magnéticas)– El ITER se encuentra en estado de construcción y el DEMO, en etapa de proyecto, se supone será el primer reactor comercial*.

Este bombardeo no solo induce radiactividad, sino que cambia las propiedades físico-químicas por las que fueron elegidos tales materiales (resistencia a las altas temperaturas, resistencia a las grandes tensiones mecánicas, etc.). En el confinamiento magnético, debido a los intensísimos campos, las estructuras estarán sometidas a enormes tensiones mecánicas y aceptarán mal las pérdidas de sus propiedades, eso por no hablar de la posible volatilización de parte de la cámara de confinamiento cada vez que se produce una disrupción, fenómeno que, a causa de inestabilidades, desvía toda la columna de plasma haciéndola impactar contra las paredes (cortocircuito), ¡y ello a temperaturas del orden de millones de grados Celsius! Por otra parte, las bobinas superconductoras, necesarias para conseguir elevados campos magnéticos, jamás se han empleado bajo el intenso flujo de neutrones que supondrá el funcionamiento del ITER (el prototipo casi comercial en fase de proyecto) y puede haber sorpresas nada agradables al respecto.

Aunque los materiales irradiados procedentes de la fusión son menos radiactivos que los correspondientes a un reactor de fisión y podrán ser manipulados sin peligro en cuestión de un par de decenas de años, los defensores de la fisión atacaron fuerte y uno de sus argumentos consistía en decir que la fusión era igual de sucia y por el contrario mucho más cara. Hace unos 35 años, Lawrence Lidsky, profesor de Ingeniería Nuclear del MIT (Massachusetts Institute of Technology), publicó un artículo atacando la fusión, en la revista de esa institución "*Technology Review*", del cual se hizo eco el *Washington Post* y terminó por dar la vuelta al mundo. Los titulares de los periódicos eran de este tenor: *"La fusión no es posible"; "Se cree que la fusión termonuclear ha llegado a su fin"; "Se considera la fusión sencilla pero inútil"; "Un veterano de la fusión declara que los problemas a los que se enfrentan son irresolubles"...* Los entusiastas de la fusión tampoco acertaban con los argumentos para rebatir el ataque.

Los razonamientos técnicos de Lidsky eran los siguientes:

a) Los costes de construcción de un reactor de fusión serán siempre mayores que los correspondientes a un reactor de fisión de la misma potencia.

b) Los problemas de ingeniería serán mayores, empezando por la necesidad de evitar el bombardeo neutrónico de los materiales estructurales que deben soportar tensiones mecánicas hasta ahora inimaginables.

c) Aunque no sean muy graves, los accidentes provocarán en un reactor de fusión paros que harán no rentable económicamente una central de este tipo.

d) Los reactores de fusión serán enormes, complicados, caros y poco fiables.

e) El uso de un reactor de fusión estará limitado a formar tándem con uno de fisión en el mejor de los escenarios.

En cualquier caso, la reacción política no se hizo esperar y Lidsky fue destituido de su cargo de director adjunto del Centro de Fusión del MIT. El encargado de rebatir sus manifestaciones fue Harold Furth, de Princeton, quien utilizó la batería de siempre. La justificación "estrella" consistió en decir: *"Si los hermanos Wright no hubieran construido un primer aparato volador torpón, nunca se hubiera desarrollado la tecnología aeronáutica actual"*, en una clara defensa de la forma de investigar en modo acierto/error. Evidente, pero también debemos reconocer que el método lo emplearon con prototipos pequeñitos. ¿Se imagina el lector a los hermanos Wright intentando hacer volar un primer modelo de la talla de un Jumbo?

De todo lo anterior se desprende que la mayor desventaja de la fusión controlada, frente a otras fuentes de energía quizás menos prometedoras pero sí más fáciles de manejar, incluida la fisión, reside en el factor económico. ¿Qué compañía arriesgará su capital en construir una planta de fusión que, en el mejor de los casos, funcionará a trompicones y producirá energía a un coste superior al de cualquier otro sistema? Primero se necesita un esfuerzo suplementario para simplificar la complejidad tecnológica de las máquinas actuales y minimizar su coste. Hoy día un reactor de fusión requiere el concurso de muchos países y, si fuera un sistema prometedor, más de una compañía eléctrica ya habría entrado en tales consorcios; no olvidemos que se postula como la energía más limpia, barata e inagotable de la historia presente y futura de la Humanidad.

Por otra parte, muchas organizaciones ecologistas ya han levantado la voz sobre la falacia de la inagotabilidad del combustible. Si la fusión del combustible mezcla Deuterio-Tritio ha sido inviable hasta hoy, más difícil será conseguir la fusión mediante la mezcla Deuterio-Deuterio. Como el Tritio es radiactivo y de corta vida media, no se encuentra en abundancia suficiente en la Naturaleza y hay que obtenerlo por bombardeo neutrónico del Litio6, por consiguiente, es el Litio el elemento que debe proveer de combustible a una planta termonuclear y su abundancia, aunque es 15 veces mayor que la del Uranio, no es ilimitada, sobre todo la del isótopo Litio6, el cual representa solo un 7 % del Litio natural. A ello se añade el problema del uso del radiactivo Tritio, susceptible de pasar al ciclo biológico del agua, y de los materiales, también radiactivos, que saldrán de las estructuras desechadas.

La fusión, por lo tanto, podría ser una gran mentira, como algunos la califican.

Supongamos que estamos en 2050 y los prototipos como el ITER han demostrado que la fusión es viable a un coste de, pongamos por caso, 8000 millones de euros por planta energética. Además, la Física del Plasma no ha proporcionado ninguna sorpresa nueva con relación a las inestabilidades, disrupciones, etc., y los reactores se han mostrado estables. El problema seguiría siendo el mismo que el actual en relación a las plantas de fisión. Una central de este estilo siempre será un generador de base, es decir, debe proporcionar energía eléctrica a la red a un ritmo constante y sin interrupciones imprevistas, pues su coste inicial hace que solamente sea rentable si está siempre operativa. Por otra parte, es muy probable que siempre se necesiten enormes cantidades de energía para reiniciar la ignición tras una parada, por consiguiente, desde el punto de vista del coste energético, una parada en fusión supone un problema mucho mayor que en fisión.

Hablemos del coste del combustible en el confinamiento magnético. Sin lugar a dudas es prácticamente despreciable frente a la inversión inicial en infraestructuras, pero esta es tan elevada, al nivel actual de conocimiento, que supera con creces los costes de otras fuentes de energía. Por esta razón muchos opinan que el tiempo y el dinero dedicado a la investigación de fusión deberían dedicarse a otras tecnologías energéticas que ya hubieran mostrado su viabilidad.

¿Es viable económicamente la fusión por láser? La respuesta, de momento, es negativa y, esta vez, aunque parezca mentira, por cuestiones de coste del combustible. Efectivamente la fabricación de los "perdigones" de combustible es hoy un proceso carísimo (véase la figura 2). El presupuesto de la Universidad de Rochester es de varios millones de dólares por 6 cápsulas de Deuterio-Tritio que se introducen en la cámara de irradiación para quemar en un año. Una máquina comercial necesitaría unas 90 000 de esas cápsulas por día. La pregunta es obvia: ¿es posible abaratar hasta extremos rentables la producción de esas cápsulas de D-T?





Figura 2: *Proyecto NIF (National Ignition Facility) del* Lawrence Livermore National Laboratory*, California.*
Izquierda*: Microcápsula de combustible ("perdigón" mezcla Deuterio-Tritio).*
Derecha*: Cámara de irradiación en cuyo centro debe implotar la microcápsula al ser iluminada simultáneamente por los haces de un láser de 500 TW. En el centro de la parte superior puede verse a un operario en labores de mantenimiento.*

Debemos pues concluir que el enfoque actual de un mundo basado únicamente en la energía de fusión no solo es poco realista, sino que en ningún caso es la panacea, ni energética, ni económica, ni política, ni social.

Rebecca Harme, del grupo de "Los Verdes" del Parlamento Europeo, solicitando en la Comisión de Industria que el esfuerzo se desvíe en otra dirección, ha asegurado: *"En los próximos cincuenta años la fusión nuclear no va a luchar contra el cambio climático ni a garantizar la seguridad de nuestro suministro energético"*. Hoy, en plena crisis económica, la opinión generalizada es que invertir 13 000 millones de euros en un proyecto a cincuenta años, del cual no tenemos asegurada la viabilidad, es descabellado. Claro que cabe preguntarse: ¿cuánto ha costado el rescate bancario, solo en España?; ¿cuál va a ser el resultado del mismo? Evidentemente se pueden hacer críticas a la fusión pero deben ser más serias y quizás provenir de ambientes más cualificados. Sin embargo, la respuesta desde el sector científico a estas críticas es siempre la misma, y siempre igual de tibia, se limita a afirmar que la seguridad intrínseca de esta forma de producir energía es suficiente pretexto como para intentar ganar la partida.

Afortunadamente hay voces discordantes más cualificadas que piensan, no que la fusión termonuclear controlada es un imposible, sino que el camino elegido, el ITER (iter es "camino" en latín), no es el correcto. Estas críticas provenientes de la comunidad científica del plasma son las más demoledoras, y no son recientes sino que incluso se remontan a los comienzos de la puesta en marcha de tokamaks como el TFTR americano y el JET europeo. El problema sobrevino cuando los demás dispositivos de confinamiento magnético fueron abandonados a favor de los tokamak, sin haber sido estudiados con la profundidad requerida ante una decisión de tal envergadura. Los primeros en caer en el olvido, pese a su simplicidad, fueron los espejos magnéticos y el laboratorio de Livermore, a comienzos de la década de los ochenta del pasado siglo, estaba a punto de inaugurar uno de esos dispositivos, también de un tamaño considerable, cuando pocas semanas antes recibió de la administración Reagan la orden de abandonar el proyecto en favor de la fusión por láser, con connotaciones más belicistas, relacionadas con el famoso proyecto conocido popularmente como "guerra de las galaxias".

En definitiva, no se han dejado madurar las ideas, no se han hecho suficientes experimentos de laboratorio a pequeña escala. Se ha cercenado la investigación básica. Algunos países como España se han lanzado a gastar dinero en construir máquinas inservibles, cuando hubiera costado mucho menos tener a los científicos y técnicos preparados con estancias en los centros donde se cocían de verdad las grandes soluciones. El propio Bob Hirsch, defensor a ultranza de las grandes máquinas tokamak a comienzos de su carrera, "cuando era un joven con prisas", –como dijo alguien–, en un discurso ante la Sociedad Nuclear Americana en 1985, atacó la idea del tokamak como inservible para los objetivos de la energía comercial de fusión. ¿Llevaría razón? Sus argumentos científicos terminaban con la idea de la inviabilidad a nivel comercial de un sistema que consistía en una complicada geometría, con anillos y más anillos abrazando una cámara de vacío a la que apenas se podía llegar ni con radiación electromagnética; *"la industria preferirá una geometría sencilla"*, afirmaba. Hirsch abogaba entonces por dispositivos más pequeños y manejables.

También hay voces que opinan que el problema reside en la falta de presupuesto. ¿Si se hubiera hecho el mismo esfuerzo por desarrollar la Física del Plasma como se ha

hecho para encontrar "el color" de los quarks, sería hoy la fusión una realidad? Es más, bastaría haber dedicado el mismo presupuesto a otros diseños que el dedicado al desarrollo del tokamak, quizás con eso hubiera sido suficiente para encontrar la solución.

Don Grove, antiguo colaborador de Spitzer, padre de los "**stellarator**" –otro diseño de confinamiento magnético para la fusión–, y a la sazón responsable del grupo de fusión de la Universidad de Princeton, se quejaba de lo que parecía evidente: cuánto menos práctico era un proyecto mayor apoyo recibía de los políticos, con tal de ganarle la carrera a algún "enemigo". Erik Storm, líder de la fusión por láser, compartía el mismo pesimismo: *"Creo que las soluciones a los problemas de la Humanidad pasan por descubrir el verdadero color de los quarks"*, decía. Claro que, al fin y al cabo, la comunidad de fusión se pasaba la vida diciendo que ya estaba a punto de encontrar la energía más segura, inagotable y barata, pero sus proyectos eran tan caros, que era más rentable escurrir las últimas gotas de petróleo de los sitios más inaccesibles.

La última crítica conocida es más feroz, si cabe, y proviene de Jean Pierre Petit, exdirector de investigación del CNRS francés y gran especialista en Física del Plasma. Petit afirma que el problema fundamental reside en el propio comportamiento del plasma de un tokamak; las disrupciones no permitirán jamás que un dispositivo comercial llegue a funcionar. Estos "apagones" del gas ionizado, de consecuencias desastrosas, están muy lejos de evitarse y pueden aparecer en cualquier momento, basta una ligera inestabilidad desencadenada por unos cuantos átomos de impurezas mezclados con el combustible.

Es posible que el problema mayor del ITER provenga del empeño de los socios en no hacer un fondo común y designar un equipo de realización del proyecto. Cada país participante quiere hacer revertir su inversión en la industria propia. En palabras de Brumfiel periodista de la revista "*Nature*" que sigue el proceso de su construcción desde hace más de 10 años: *"es como si uno pidiera bajo catálogo una serie de chapas y cientos de miles de tuercas y tornillos para construir en el jardín de su casa un Boeing 747"*. Pero hay un par de detalles que todos olvidan: 1) en plena crisis no es difícil que alguna de las empresas adjudicatarias de parte del proyecto quiebre a la mitad o casi al final del camino, con el dinero gastado y la pieza a medio hacer; y 2) tras los muchos ajustes, hay piezas que nunca han sido especificadas en los catálogos, posiblemente hasta la hora de ensamblar el conjunto no se sepa cuáles son. Efectivamente, algunos países, como la India, empiezan a tener problemas para cumplir sus compromisos o para repartir los contratos entre las empresas subcontratantes.

En resumen, podemos afirmar que, por mucho tiempo, la fusión no será la protagonista del futuro energético de la Humanidad.

Julio Gutiérrez Muñoz
Doctor en Física
Catedrático de Universidad de Física Atómica, Molecular y Nuclear, jubilado.

Capitulo 57
¿Cómo funciona un microscopio electrónico? ¿Existen microscopios basados en otras partículas fundamentales?

(Por Francisco J. Terán)

Desde finales del siglo XVI, la microscopía se ha venido desarrollando como una rama de la ciencia con el fin de proporcionar tecnologías que faciliten la observación y el estudio de objetos demasiado pequeños de ser percibidos a simple vista. A día de hoy, no está del todo claro si la primera "*mirada*" de la microscopía fue puesta en el firmamento para estudiar esas *"pequeñas luces"* que allí brillan denominadas *"astros"* -Galileo Galilei no fue el pionero en la creación del telescopio pero si quien más lo mejoró y publicitó (1)- o en los microorganismos -Antonie van Leeuwenhoek fue pionero en microbiología por sus mejoras en el microscopio (2)-.La microscopía óptica, basada en el uso de luz, lentes y otros componentes ópticos, fue la primera en desarrollarse. Con ella se asentaron buena parte de los fundamentos generales para otras microscopías que también emplean ondas para sondear la materia en la escala submicrométrica. Así, se abrieron las puertas a grandes avances en distintas áreas de la ciencia, desde el estudio de la célula a las propiedades de la luz o la materia. Esto último fue clave para el descubrimiento de ciertas leyes de la física a partir de las cuales han surgido microscopías como la electrónica, permitiendo la visualización y manipulación de objetos en la escala del nanómetro (1 nm = 10^{-9} m). A principios del siglo XX, de Broglie (3) enunció el carácter ondulatorio del electrón. La *"óptica electrónica"* superó las limitaciones de la microscopía óptica en el rango visible (400-700 nm) gracias a que de Broglie relacionó la longitud de onda del electrón con su energía cinética ($\lambda_{dB}=h/p$, donde h es la constante de Planck y p el módulo del momento del electrón). De esta forma, acelerando al electrón (i.e. aumentando p) se puede reducir λ_{dB} varios órdenes de magnitud. Consecuentemente, se pueden lograr aumentos del orden de x10^6 con una resolución espacial del orden del angstrom (1 Å= 10^{-10} m), valores muy alejados de los que ofrece la microscopía óptica tanto en aumentos (~x10^3) como en resolución espacial (~200 nm).

A continuación describiremos los fundamentos básicos de las modalidades más relevantes de la microscopía electrónica. Para finalizar, haremos mención a otras microscopías basadas en el carácter ondulatorio de otras partículas.

57.1 Así funciona un microscopio electrónico

En la actualidad, la microscopía electrónica es una técnica ampliamente extendida para caracterizar formas, dimensiones, superficies y composición química de la materia en la escala del nanómetro. A pesar de compartir fundamentos ondulatorios teóricos con la microscopía óptica, la electrónica implica una tecnología mucho más compleja. La microscopía electrónica posee principalmente tres modalidades:

microscopía electrónica de barrido (en inglés Scanning Electron Microscope, SEM), transmisión (en inglés, Transmission Electron Microscopy, TEM), o de efecto túnel (Scaning Tunelling Microscope, STM). De estas, se derivan a su vez otras modalidades subyacentes como la microscopía electrónica de transmisión en barrido (en inglés, scanning transmission electron microscope) o la microscopía electrónica de reflexión (en inglés, reflection electron microscope).

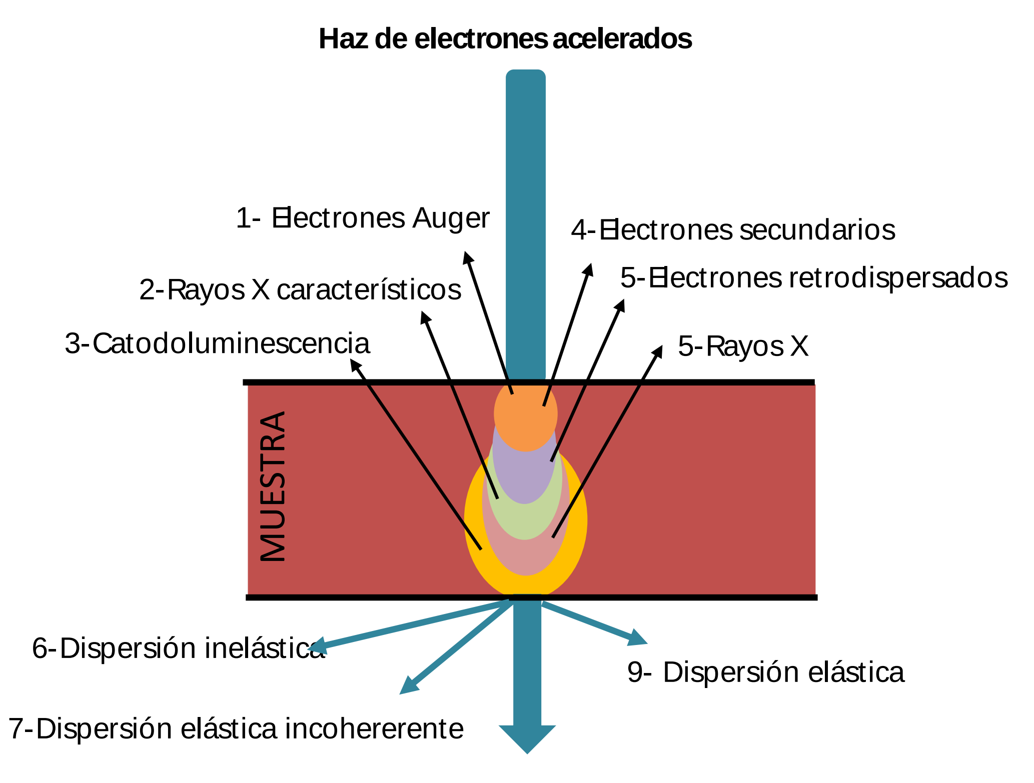
Históricamente, el primer microscopio electrónico se desarrolló en los años 30 del siglo XX. Ernst Ruska desarrollo en Alemania el TEM cuyo diseño original mantiene en la actualidad las partes principales. En su origen, el desarrollo del TEM se benefició además de la demostración por Ruska que un electroimán actúa como una lente electrónica. Ambos avances, la capacidad de controlar la energía del electrón (i.e., λ_{dB}), y de generar un haz de electrones de trayectoria variable, hicieron posible la construcción el primer TEM en 1933. Siemens inicio su comercialización en 1939. De esta manera se inició el estudio de objetos cuyas dimensiones están en la escala del nm, que no ha parado hasta hoy. En paralelo al desarrollo del TEM, Manfred von Ardenne inventa en 1937 el SEM apoyándose en descubrimientos de Max Knoll, quien también participo en la invención del TEM. El funcionamiento del SEM se basa en el barrido de una región espacial mientras incide un haz de electrones. Así, se puede estudiar superficies en base al análisis de la interacción de los electrones con la superficie, por ejemplo, analizando los electrones secundarios emitidos por los átomos de la superficie. Finalmente, en 1981 Gerd Binnig y Heinrich Rohrer desarrollan el STM, un microscopio electrónico basado en el efecto túnel, un fenómeno cuántico que permite a un electrón atravesar una barrera de potencial mayor que su energía cinética. Es decir, permitir el paso de una corriente de electrones desde el último átomo de una punta metálica a otro ubicado sobre la superficie de la muestra (metálica o semiconductora). Esto es posible cuando la distancia entre la punta y la muestra es de unos pocos Å y se aplica una tensión generalmente inferior a tres voltios entre la punta y la muestra (4). De esta forma se pueden obtener imágenes sin precedentes de átomos individuales en superficies con una resolución inferior a 1 Å, tanto lateral como verticalmente. La clave de la sensibilidad STM reside en que la corriente túnel de electrones varía un orden de magnitud por Å de distancia punta a muestra. Pequeñas variaciones topográficas se traducen en grandes cambios de corriente. Debido a la naturaleza del propio efecto túnel, la corriente de electrones que *"tunelean"* desde la punta hasta los átomos de la superficie sondea orbitales electrónicos de valencia del átomo. Por tanto, la microscopía STM también da información de los orbitales electrónicos, permitiendo describir la estructura electrónica de las superficies. Pero la característica más fascinante del STM es la manipulación de la posición de átomos individuales, variando su ubicación espacial a voluntad (5). Uno de los componentes básicos del STM a los que debe su éxito y avance son los soportes piezoeléctricos sobre los que se montan las puntas. Estos soportes permiten mover la punta con total precisión distancias < 1 Å y mantenerla días con plena fiabilidad. Esta microscopía despierta un interés exclusivamente académico, para estudiar las propiedades de la materia relacionadas con la morfología de superficies, moléculas individuales, propiedades eléctricas y magnéticas de moléculas y átomos individuales, afinidades y enlaces electrónicos de átomos y moléculas individuales, síntesis de nuevos materiales, propiedades electrónicas de materiales topológicos, entre otros temas.

Como se señalaba anteriormente, las microscopías electrónicas (salvo la STM) se benefician de la modulación de λ_{dB} a través de aumentar el momento del electrón para alcanzar extraordinarios valores de resolución espacial y aumentos. Pero es sin

duda, el análisis de los distintos procesos de interacción de un electrón con los átomos de la muestra en estudio lo que marca la mayor diferencia con la microscopía óptica (ver figura 1). Así, las microscopías SEM y TEM son las técnicas de caracterización estructural y composicional de nanomateriales más empleadas tanto por científicos como por ingenieros. Ambas tienen similitudes y diferencias en su tecnología. Las similitudes son en cuanto al uso de fuentes de electrones para generar el haz, lentes electromagnéticas para focalizarlo en una determinada ubicación espacial de la muestra, y sistemas de vacío (vacío más alto para TEM) para minimizar la desviación del haz a fin de mejorar la nitidez de la imagen. Por un lado, la generación del haz de electrones emplea habitualmente filamentos de tungsteno o dispositivos de efecto campo con el fin de crear haces de electrones coherentes que mejoren el contraste de la imagen. Tras la generación del haz de electrones en la fuente, se les somete a un alto voltaje con respecto a la muestra para dotarles de un momento *p* definido. Estos valores de voltaje suelen ser de unos ~100 kV para un TEM estándar hasta 200 kV para un TEM de alta resolución. En el caso del SEM, el voltaje aplicado es inferior a los 40 kV. Por otro lado, las lentes electromagnéticas se basan en la fuerza que ejerce un campo magnético sobre una carga eléctrica que se mueve con una determinada energía cinética $\vec{F} = q\vec{v}\text{x}\vec{B}$. También, los campos eléctricos pueden deflectar un ángulo fijo la trayectoria de un electrón. La aplicación de una fuerza magnética y/o eléctrica permite desplazar lateralmente la trayectoria de un haz de electrones para focalizarlo en una posición especial determinada. Para ello, se requiere que la muestra esté en un vacío cuya presión típica oscile entre 10^{-4} y 10^{-8} Pa. La necesidad de vacío en la cámara de la muestra limita consecuentemente tanto el modo de introducir las muestras como su preparación. Así por ejemplo, las muestras biológicas (que poseen un gran contenido en agua) requieren ser inmovilizadas en resinas plásticas (fijación química, intercambio de agua e inclusión en resina) o ser preservadas en frío (crío-fijación) para no alterar las condiciones de vacío.

Por el contrario, los SEM y TEM poseen grandes diferencian en cuanto al tipo de muestras que pueden estudiar y a su preparación. Por un lado, las muestras SEM han de ser conductoras o de lo contrario ser metalizadas para evitar acumulación de electrones (carga) lo cual altera la calidad de la imagen. Por otro lado, las muestras TEM pueden ser cualquier tipo de muestras amorfas, cristalinas, biológicas, etc. Su única condición es que el espesor ha de ser ultrafino (inferior a 500 nm) para permitir la transmisión de electrones a través. También, las técnicas de visualización de imágenes son distintas para el SEM y TEM ya que analizan diferentes procesos de interacción electrón-materia (ver figura 1), sean relacionados con la transmisión (TEM) o con la difusión (SEM) del electrón sobre la muestra. Así, el TEM proporciona imágenes de mayor resolución (hasta 3 Å cuando la aberración esférica de tercer orden está corregida (6)) que son una proyección bidimensional generada por la superposición de las funciones de ondas de los electrones que atraviesan la muestra. Sin embargo un SEM proporciona imágenes tridimensionales de la superficie pero con una menor resolución espacial (~10 nm). A los correspondientes sistemas de visualización SEM y TEM, se les puede complementar con otro tipo de análisis (rayos X, catodoluminescencia, pérdidas de energía del electrón, procesos Auger, electrones secundarios, etc.…) que proporcionan valiosa información sobre la muestra (composición química, número atómico, estado electrónico). En la actualidad, la gran revolución ha llegado de la mano de los detectores directos de electrones, que han permitido la toma de datos con mucha menor intensidad del haz y mejor contraste. Tradicionalmente, los histogramas o imágenes de SEM y TEM se han venido haciendo en blanco y negro, porque la técnica elimina los matices cromáticos. Recientemente se ha logrado la manera de colorear

dichas imágenes empleando un detector que captura los electrones que devuelven los iones en color (7). También, por el avance informático que permite controlar y analizar imágenes en las que se combinan imágenes morfológicas con *"mapeados"* de la muestra en relación a distintos análisis (por ejemplo, presencia de un elemento químico). La llegada de las cámaras CCD y la mejora en la informática permitió la automatización de los procesos de visualización de imágenes y de control de los componentes del microscopio.



Interacción electrón-materia	Información extraíble	Interacción electrón-materia	Información extraíble
1	Composición atómica de la superficie	6	Composición y estados ligados
2	Composición atómica	7	Composición atómica y estructural
3	Estados electrónicos	8	Morfología
4	Topografía	9	Análisis estructural e imágenes de alta resolución
5	Número atómico y diferencias de la fase de la función de onda del electrón		

Figura 1: *Representación esquemática de las distintas modalidades de interacción de un haz de electrones y la materia.*

57.2 Otras microscopías basadas en otras partículas fundamentales

Como hemos mencionado anteriormente, la microscopía electrónica se basa en el carácter ondulatorio del electrón para el análisis de la materia. No existen otras microscopías alternativas basadas en el carácter ondulatorio de otras partículas fundamentales con carga por la dificultad en las condiciones de su generación y la imposibilidad de desarrollar fuentes (estas partículas se generan típicamente en instalaciones de alta energía). Sin embargo, si ha sido posible desarrollar microscopios que emplean haces de partículas más pesadas, como son los iones de gases nobles (He, Ne, Ar), moléculas de hidrogeno (H_2) o mezclas de ambos. La invención de la microscopía de iones de campo (en inglés, Field Ion Microscopy, FIM) por Erwin Wilhelm Müller en 1951 permite analizar superficies con una metodología similar a la del SEM (análisis de electrones secundarios). Sin embargo, su resolución espacial (< 3 Å) es dos órdenes de magnitud superior gracias a que la longitud de onda de los iones es más corta que la de los electrones. Otra de sus ventajas respecto al SEM es que la muestra no requiere ser un conductor eléctrico. Las generalidades del funcionamiento de un FIM son similares a las de un SEM, variando la fuente de ondas (iones), la refrigeración de muestra (a temperaturas entre 10 y 150 K) que está colocada en un soporte redondeado (de radio entre 10 y 100 nm) y el sistema de visualización de la superficie de la muestra que incluye placas con microcanales para amplificar la señal de los iones dispersados. La versión más avanzada del FIM es la que emplea iones de He comercializado desde 2007 (8) debido a que los iones de He no alteran la muestra en estudio por la ligera masa del He. Finalmente, mencionar que si bien la microscopía óptica se basa en luz visible, luz de alta energía como son los Rayos X (con longitudes de onda 2-4 nm, cien veces más pequeña que la luz visible) ofrece una resolución espacial del orden de los 30 nm. En 1950, Sterling Newberry desarrolló el primer microscopio de rayos X que fue comercializado por General Electrics. A día de hoy, se han realizado versiones que permiten un barrido de la muestra (en inglés, Scanning X-Ray microscope) y se benefician de las cámaras CCD y el avance de la informática para el control del microscopio, y el análisis de imágenes.

Notas:

1. Gribbin, John (2006). *Historia de la ciencia. 1543-2001*. Crítica, p. 82.
2. Finlay BJ1, Esteban GF. Int Microbiol. 2001 Sep;4(3):125-33.
3. Louis Victor de Broglie, RECHERCHES SUR LA THÉORIE DES QUANTA (Ann. de Phys., 10e série, t. III (Janvier-Février 1925).
4. Gerd Binnig and Heinrich Rohrer" *Scanning tunneling microscopy—from birth to adolescence*" Rev. Mod. Phys. 59, 615 – Published 1 July 1987.
5. Famosa es la imagen de STM con el texto "IBM" que escribieron con átomos individuales los descubridores del STM mientras trabajaban en el Centro de Investigación IBM en Zürich (Suiza).
6. La aberración esférica es el óptico por excelencia en la limitación de la resolución espacial del TEM y SEM.
7. Adams, Stephen R. et al., Cell Chemical Biology, Volume 23, Issue 11, 1417-1427
8. Orion Helium Ion Microscope, ZEISS Group

Francisco José Terán Garcinuño
Doctor en Ciencias Físicas
Investigador en iMdea Nanociencia

Capítulo 58
¿Qué es la entropía y qué relación tiene con el tiempo?
(Por Pablo López Tarifa)

"La energía de un sistema ni se crea ni se destruye, solo se transforma".

Así nos enunciaban el primer principio de la termodinámica en nuestra primaria básica. Reformulado de manera más precisa, este principio establece que la energía de un sistema aislado permanecerá constante, y que todo flujo de energía que haya en su interior será debido al intercambio de calor o en trabajo entre sus diferentes partes, haciendo que la suma de ambas magnitudes sea siempre constante al valor inicial.

Independientemente de como se enuncie, este principio constituye una piedra funda-mental de la física y su cumplimento se impone hasta en el más simple de los modelos teóricos. De una forma pragmática, nos permite calcular el rendimiento de un dispositivo electrónico o saber cuánto combustible va a necesitar el motor de un coche para moverse. Se puede decir que el primer principio de la termodinámica establece cómo los cuerpos intercambian energía, ya sea en forma de calor o trabajo, pero, sin embargo, no especifica nada de en qué sentido lo hacen. La entropía, o más específicamente el segundo principio de la termodinámica, es el encargado de señalar la dirección en la que la energía fluye, estableciendo una relación especial con el tiempo.

58.1 La entropía y su interpretación estadística.

En nuestra vida cotidiana podemos experimentar el flujo de energía. Por ejemplo: vemos que el agua de los ríos cae montaña abajo, que la batería de nuestro móvil se descarga durante el día o que los hielos puestos en una bebida caliente se deshacen con el tiempo. Si pensamos en términos de energía, en todos estos casos hay una diferencia en su concentración, es decir, una parte concreta del sistema posee más (o menos) energía que su entorno: El agua de los ríos cuando está cerca de la montaña tiene alta energía gravitacional, al igual que la batería de nuestro móvil recién cargada tiene alta energía química, o los hielos recién sacados de la nevera tienen muy baja energía térmica. Esta alta (o baja) concentración de energía tiende a disiparse con el tiempo, es algo que sucede de forma espontánea y parece que es una tendencia que cumplen todos los objetos de nuestro universo. Se puede decir por tanto que la energía tiende a la uniformidad de forma natural.

El término de "entropía" fue precisamente acuñado por el físico alemán Rudolf J. E. Clausius en 1850, como medida del grado de uniformidad energética de un sistema. Cuanto más uniforme sea, mayor será el valor de su entropía. Si la energía está distribuida de manera perfectamente uniforme, la entropía obtendrá su valor máximo. Realizando simples experimentos como poner objetos calientes en contacto con otros

fríos, o igualando niveles de depósitos de agua a diferentes alturas, mostró como cualquier diferencia de energía dentro de un sistema tiende siempre a igualarse por sí sola, estableciendo un sentido unidireccional para el flujo de energía: el calor fluía siempre del objeto caliente al frío, al igual que el agua caía del depósito situado a mayor altura al de menor. El proceso contrario, que el calor fluyera del objeto frío al caliente o que el agua fluyera de menor a mayor altura, nunca ocurría de forma espontánea. Aunque por simples que parezcan estas conclusiones no son tan obvias, ya que si se produjeran los sucesos opuestos, las leyes de conservación de la energía o la de la conservación de momento lineal se seguirían cumpliendo. Como corolario a su estudio Clausius estableció que la entropía de un sistema aumenta siempre con el tiempo, esencia de lo que hoy conocemos como el segundo principio de la termodinámica.

La relación del flujo de energía entre sistemas con diferentes concentraciones continuó estudiándose intensamente durante la segunda mitad del siglo XIX, especialmente en su relación con la energía térmica. En 1877, el físico austriaco Ludwig Boltzmann encontró la manera de expresar matemáticamente el concepto de entropía, usando lo que hoy se conoce como su interpretación estadística:

$$S = k_B \ln W$$

En la fórmula, S representa la entropía, k_B es la llamada constante de Boltzmann (1.38064852×10^{-23} m^2kg/s^2K) y ln W es el logaritmo neperiano de la probabilidad termodinámica, W. Para ilustrarla, tomaré el ejemplo original de Asimov: Imaginemos que a nueve personas ordenadas en un cuadrado de tres columnas de tres, separadas las filas y columnas uniformemente, se les piden que den un paso en cualquiera de las cuatro direcciones del espacio (hacia adelante, hacia atrás, hacia un lado o hacia el otro). Las nuevas posibles configuraciones van a ser un total de:

4 x 4 x 4 x 4 x 4 x 4 x 4 x 4 x 4= 262.144

que estadísticamente se las denomina como microestados del sistema. Atendiendo a un segundo criterio, llamado macroestado, podemos clasificar los microestados en “ordenados” si las nueve personas están en una formación equidistante, simétrica, fácil de visualizar y fácil de describir (como la inicial); o por el contrario “desordenados” si no hay equidistancia alguna. Usando la calculadora podemos calcular el valor de la W de la fórmula de Boltzmann, es decir, la probabilidad de que un macroestado tenga lugar. Echando cuentas, vemos que son tan solo 4 microestados los que pertenecen al macroestado ordenado, es decir 1 posibilidad entre 65.536 de que las nueve personas decidan moverse todas hacia un mismo lado. Sin embargo, los 262.140 microestados restantes pertenecen al macroestado desordenado, de mayor entropía. Parece inevitable entonces que a cada paso que se dé, el sistema formado por las nueve personas acabe inexorablemente en un estado desordenado. Incluso si, por casualidad, para el primer paso todas las personas se moviesen en bloque hacia un lado, es casi seguro que el siguiente paso romperá la formación.

El mismo argumento se puede utilizar si hablamos en vez de personas, de átomos, y en vez posiciones, por ejemplo, de velocidades atómicas, estrechamente relacionadas con la temperatura y la energía térmica de un sistema. La interpretación de Boltzmann nos dice que la energía tiende a la uniformidad porque estadísticamente es el resultado más probable.

Aunque interpretaciones más modernas sugieren una definición de entropía en función de la información contenida en el sistema (por ejemplo, cuántas palabras son necesarias usar para describir unívocamente su estado), la interpretación estadística es perfectamente válida, siendo vigente hoy en día y con extendido uso. De hecho, sustituyendo la probabilidad termodinámica por las llamadas funciones de partición, uno puede determinar la entropía de, por ejemplo, un cambio de fase o a la formación de un determinado compuesto químico, datos que nos ayudan a predecir de antemano si una reacción química podría ocurrir de forma espontánea o no.

58.2 La entropía y el tiempo.

En nuestra experiencia, no tenemos duda de que el tiempo fluye y que éste lo hace en una dirección particular: Sabemos que todos lo días el Sol sale por el este y pone por el oeste, que las horas pasan y que los días que nos separan de un evento en particular, terminarán por llegar. Tenemos el recuerdo de que éramos jóvenes en el pasado y la seguridad de que seremos viejos en el futuro. Además, nuestro cerebro puede recordar eventos pasados, pero nunca *recuerda* eventos del futuro.

A esta diferencia entre pasado y futuro los físicos la llaman asimetría temporal, y como hemos visto se puede verificar fácilmente en todos los procesos que nos rodean. Por ejemplo, imaginemos que grabamos en una cámara cinematográfica la caída de un huevo sobre el suelo. Imaginemos también que la cinta es cortada en los diferentes fotogramas y éstos son barajados aleatoriamente. Si se nos pide reordenarlos de nuevo, no cabrá ninguna duda de que situaremos el fotograma con el huevo intacto el primero de todos, haciendo un orden cronológico descendente situando el fotograma con el huevo roto el último de todos. Podremos decir que este es el sentido natural del suceso, y que el contrario, que los fragmentos del huevo vuelvan a juntarse para formar el huevo intacto nuevamente, nunca tendrá lugar. Sin embargo, por extraño que parezca, en las leyes de la física no prevalece un proceso sobre el otro. Desde un punto de vista teórico ambos podrían suceder, solo bastaría con invertir las velocidades de cada uno de los pedazos del huevo, y el proceso inverso tendría lugar. Entonces, ¿por qué sucede el proceso hacia adelante en el tiempo y no el inverso?

La respuesta tiene estrecha relación con el hecho de que en las leyes de la física no existe una diferencia intrínseca entre pasado y futuro. Es más, el sentido del fluir del tiempo parece desdibujarse cuando se consideran sistemas mucho más sencillos, con menos partículas. Como señalaba Asimov en su escrito original, un electrón curvándose en determinada dirección con el tiempo marchando hacia adelante podría ser también un positrón curvándose en la misma dirección pero con el tiempo marchando hacia atrás. En estas circunstancias, nos resulta difícil discernir entre pasado y futuro, saber si la película que estamos viendo va hacia adelante o hacia atrás. El porqué de esta confusión hay que buscarlo en la entropía que, como vimos en el apartado anterior, siempre aumenta para un sistema complejo. Nosotros aceptamos este aumento como señal que todo se desarrolla *normalmente* y que nos movemos hacia adelante en el tiempo. Por oposición, en un sistema con un número reducido de partículas simples, cualquier movimiento de una de ellas lleva asociado un cambio de entropía mínimo, casi nulo. En estas circunstancias es cuando nos quedamos cual marino sin brújula, sin cambio de entropía que nos indique hacia dónde marcha el tiempo, haciéndonos muy confuso el discernir si el proceso va hacia delante o hacia atrás. Por este motivo a la entropía se le suele denominar comúnmente como la flecha del tiempo. Sin embargo, cabe diferenciar

entre sentido y flujo del tiempo. Que la entropía de un sistema aumente con el tiempo no significa necesariamente que el tiempo pase, al igual que una brújula señala hacia el norte, no significa que nos movamos hacia él. La entropía señala una asimetría del tiempo, o más específicamente una asimetría de los estados de los sistemas en el tiempo. No dice nada de que el tiempo fluya. De hecho, ¿cuánto tiempo pasa en un segundo? La respuesta un segundo por segundo carece de sentido.

Muchos científicos piensan que efectivamente la sensación de fluir del tiempo no es más que una ilusión de nuestro cerebro. Si esto fuera cierto, imaginemos que llegue el día en el que médicos, antropólogos o psicólogos lleguen a localizar qué parte del cerebro es la encargada de proporcionarla. Puede que llegue el día en el que nos podamos operar de nuestro *sentido del fluir del tiempo*, haciendo nuestro pasado presente, o borrando nuestras preocupaciones por el futuro.

Bibliografía:
[1] "The physics of time asymmetry". Paul Davies. University of California Press, 1974.
[2] "The Quark and the Jaguar: Adventures in the Simple and the Complex". Gell-Mann Murray. St. Martin's Griffin; Ill edition, 1995.
[3] "Filosofía del tiempo". George Musser. Investigación y ciencia, ISSN 0210-136X, Nº 314, 2002 (Ejemplar dedicado a: El tiempo), págs. 14-15.
[4] "About Time: Einstein's Unfinished Revolution". Paul Davies. Simon & Schuster; 1st edition, 1996.

Pablo López Tarifa
Doctor Química
Investigador postdoctoral en la Vrije Universiteit, Amsterdam

Capítulo 59
Si el universo está en expansión ¿cómo fue al principio?

(Por Juan García-Bellido Capdevila)

59.1 Hubble y la expansión del universo.

Cuando Albert Einstein formuló la teoría de la relatividad general, hace ahora un siglo, no podía imaginar que nuestra visión del universo iba a cambiar de manera tan drástica. Unos años después, Vesto Slipher y Edwin Hubble mostraban que los objetos nebulosos que ahora llamamos galaxias se alejaban de nosotros a enormes velocidades. George Lemaitre, usando las ecuaciones de Einstein y Alexander Friedmann interpretó correctamente dichos desplazamientos como debidos a la expansión del universo. Esto cautivó la imaginación de los cosmólogos y abrió las puertas a la exploración de un posible origen del universo: si el universo está hoy en expansión, eso es que en el pasado tuvo que estar más concentrado y por tanto más caliente. Por otra parte, si el universo tuvo un origen, entonces tenemos acceso a solo una región finita del universo (que llamamos universo observable) que es aquella de la cual nos llega luz y cuyo tamaño es aproximadamente la edad del universo por la velocidad de la luz. Hoy en día nos llega luz de los confines del universo, en concreto del momento en que el plasma caliente primigenio se enfrió por la expansión lo suficiente para que se formaran los primeros átomos. Esta luz la detectamos en el rango de microondas y constituye la imagen más antigua del universo.

Ahora bien, si el universo está en expansión ¿qué agente actuó como disparador? La respuesta curiosamente estaba implícita en las propiedades del fondo de radiación de microondas. La luz de este fondo tiene un espectro de energía que corresponde a una única temperatura, cualquiera que sea la dirección en la que miremos. Según las ecuaciones de Einstein, esto no debería ser posible ya que fué emitida hace miles de millones de años, cuando el universo era mucho más pequeño, y regiones que hoy en día estan separadas por más de un grado en el cielo, entonces no estaban en contacto causal, luego no podrían haberse puesto de acuerdo en emitir fotones con la misma temperatura. A esas regiones causales se les conoce como horizontes, y son una característica ineludible de la estructura del espacio-tiempo.

59.2 Cosmología inflacionaria.

La solución vino de la mano de la física de altas energías, que postuló un periodo de expansión acelerada anterior a la formación de los primeros elementos, llamado inflación. De manera que regiones que estaban en contacto causal durante inflación dejaron de estarlo debido a la rápida expansión, que las sacó fuera del horizonte. Más tarde, cuando la inflación termina y se genera toda la materia, la expansión decelerada hace que vuelvan a entrar en el horizonte. Luego la razón por la

que vemos regiones separadas en el cielo con la misma temperatura es una consecuencia directa de ese proceso inflacionario, que dio el pistoletazo de salida a la expansión del universo. La dinámica responsable de la inflación aún no se conoce. La teoría postula un campo escalar efectivo similar al de Higgs cuya densidad de energía acelera el universo. El origen de este proceso inflacionario bien podría ser una fluctuación cuántica del propio espacio-tiempo, a escalas de energías próximas a las de la gravedad cuántica, caracterizadas por la masa de Planck, diecinueve órdenes de magnitud mayor que la masa del protón y muy lejos de poder ser explorada con los actuales aceleradores de partículas.

59.3 Fluctuaciones cuánticas y formación de galaxias.

La enorme densidad de energía responsable del proceso inflacionario hizo que el universo se expandiera al menos treinta órdenes de magnitud, hasta tamaños del órden de un centímetro. Esta fenomenal expansión allanó cualquier grumo que pudiera tener el universo primordial y lo hizo extraordinariamente plano (esto es, Euclídeo). Lo interesante es que las propias fluctuaciones cuánticas del campo del inflatón perturbaron la métrica del espacio-tiempo, dejando su huella en forma de ondas, similar a las que deja la marea en la arena cuando se retira. Al final de la inflación, la enorme densidad de energía se transformó de forma explosiva (de ahí el nombre de “Big Bang”) en otras partículas de materia y energía que llenaron esas huellas. Cuando observamos el fondo de radiación de microondas se pueden ver los grumos de materia, en los cuales se dispersan los fotones antes de viajar hacia nosotros, como pequeñas desviaciones respecto a la temperatura común del fondo. Estas pequeñas perturbaciones de materia fueron creciendo conforme se expandía el universo, hasta formar las galaxias y cúmulos de galaxias. Las propiedades estadísticas de las fluctuaciones en la temperatura del fondo de microondas y las perturbaciones en la densidad de galaxias concuerdan entre sí y están de acuerdo con las predichas por el paradigma de inflación. Sin embargo, estas fluctuaciones no nos permiten todavía medir la escala de energía a la que este proceso ocurrió. Para ello es necesario tener en cuenta que, además de grumos de densidad, la inflación genera ondas gravitacionales, cuya amplitud es proporcional a la escala de inflación. Si en un futuro próximo detectamos las huellas dejadas por las ondas gravitacionales en la polarización del fondo de radiación, podremos determinar dicha escala de energía. Hay varios experimentos tomando datos en estos momentos y varios satélites propuestos, y su detección abriría la exploración de nueva física a gran escala de energía, inalcanzable por los actuales y futuros aceleradores de partículas.

59.4 El recalentamiento después de Inflación: el Big Bang.

Como hemos descrito antes, al final de inflación se genera toda la materia y energía en un proceso explosivo asociado a la desintegración del campo del inflatón, que llamamos el Big Bang. El proceso es tan violento que debió dejar un fondo de ondas gravitacionales que eventualmente seremos capaces de detectar, aunque su frecuencia típica está muy lejos del rango de sensitividad de los detectores actuales como LIGO. Antes del Big Bang, la enorme expansión inflacionaria del universo diluyó toda la materia que pudiera haber habido antes, luego solo quedaba el campo homogéneo del inflatón, es decir, la entropía del universo era aproximadamente cero. Con la fenomenal conversión de energía en el Big Bang se formaron de golpe trillones de trillones de partículas (concretamente diez elevado a noventa partículas), y por tanto

una enorme entropía. Desde ese momento, el univeso se expande lentamente de forma cuasiadiabática, sin producción de entropía. En termodinámica cuántica hay un límite (descubierto por Bekenstein) a la cantidad de entropía que puede almacenar una región finita del espacio. La máxima entropía se alcanza cuando todo el sistema en dicha región llega al equilibrio termodinámico. A pesar de la enorme producción de partículas y entropía al final de inflación, el universo tenía un tamaño tan grande que la máxima entropía que podía contener era aún mucho mayor que la generada en el Big Bang. Eso ha permitido que en la evolución posterior del universo sea posible tener procesos que creen estructuras ordenadas que disminuyen la entropía, a costa de la entropía total del universo, que aumenta. Hoy en día los procesos de formación de galaxias por colapso gravitacional son capaces de crear orden porque aún nos queda un potencial enorme de aumento de entropía antes de llegar al equilibrio termodinámico. Esto explica, entre otras cosas, que sea posible que se dé la vida en nuestro universo, pues crea orden a costa de la entropía total, algo que obsesionaba a Lord Kelvin y a Ludwig Boltzmann a finales del s.XIX, que no entendían por qué no habíamos llegado ya a la muerte térmica en un universo eterno e infinito. Luego inflación no solo explica por qué es plano y homogéneo nuestro universo, sino que da el pisoletazo de salida del Big Bang y determina las condiciones iniciales de baja entropía que permiten la vida. Esta propiedad de inflación no suele discutirse, pero resulta esencial para comprender nuestra existencia en el universo.

59.5 Metauniverso: la estructura a muy gran escala del universo.

Nosotros observamos una región finita del universo, aquella de la cual nos llegan fotones y ondas gravitacionales desde el origen del universo. Esta región es una esfera centrada en la Tierra, con radio aproximado de cuarenta y seis mil millones de años luz, y constituye un horizonte físico. Puede parecer inmenso, inabarcable, pero esa región no comprende todo el universo. Aunque no nos llegue luz, sabemos que hay espacio-tiempo más allá de este horizonte ya que observamos que nuestro universo observable es aproximadamente euclídeo, con una curvatura espacial menor del 0.1%, luego al menos el espacio se extiende unas mil veces mas allá del universo observable.

En el contexto de inflación, el universo es esencialmente infinito, nuestra región del universo surgió hace trece mil ochocientos millones de años, pasó por un periodo inflacionario y más tarde decayó generando todo lo que vemos, pero hay regiones del universo muy alejadas de nosotros que podrían estar aún sufriendo dicho proceso de expansión acelerada. Es más, según el paradigma de inflación, dichas regiones podrían ser muy distintas de la nuestra, incluso con distintas constantes fundamentales como la masa del neutrino o la constante de interacción fuerte o el contenido de materia oscura. Hablamos por tanto de un “metauniverso”, donde las características locales no tienen por qué ser universales. Es cierto que esta estructura a muy gran escala de universo está tan lejos que no tenemos ninguna evidencia de su existencia, luego es más especulativo que ninguna de las otras predicciones de inflación.

59.6 El futuro de la expansión del universo.

En estos momentos conocemos con bastante precisión el contenido de materia y energía del universo observable. Sabemos que además de materia similar a aquella de la cual estamos hechos, hay una enorme cantidad de materia oscura, posiblemente

agujeros negros primordiales, generados cuando el universo tenía menos de un segundo. También hay una pequeñísima densidad de energía (así llamada oscura) responsable de la aceleración actual del universo, que por el momento no somos capaces de distinguirla de una constante cosmológica. Pues bien, el futuro de la expansión del universo depende muy sensiblemente de la naturaleza de dicha energía oscura. Si es una constante fundamental, entonces el universo se seguirá expandiendo de forma acelerada y diluyéndose hasta que cada galaxia terminará aislada, en lo que se conoce con el Big Chill. Alternativamente, si la energía oscura decae en materia, como ocurrió al final de inflación, podría terminar en un Big Crunch. Finalmente, si la energía oscura crece con el tiempo, entonces la aceleración se hace cada vez mayor e incluso la estructura del espacio-tiempo sufre un desgarro y terminamos con un Big Rip. Ninguna de las tres alternativas son atractivas, pero su realización dependerá del contenido de materia y energía del universo observable. Por ello se están estudiando enormes cantidades de galaxias, en grandes regiones del espacio, para poder determinar con mayor precisión cuál es la naturaleza de la energía oscura y por tanto el destino del universo. Creo que el esfuerzo bien vale la pena.

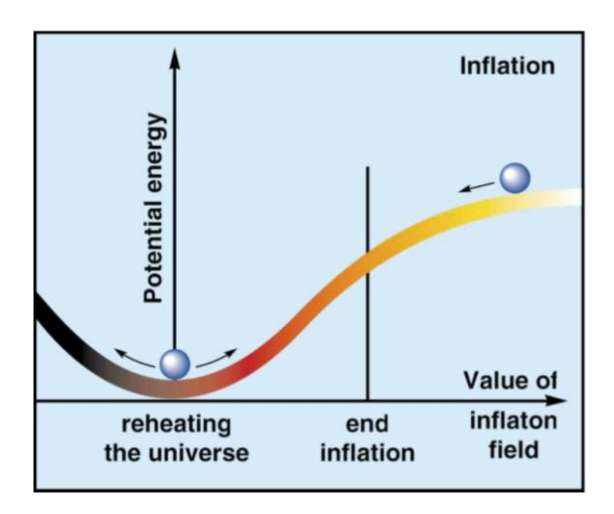


Figure Caption: *El proceso inflacionario se describe como una canica (cuya posición determina el valor del campo del inflatón) cayendo en un cuenco (cuya altura determina la densidad de energía del campo). La enorme energía aproximadamente constante del inflatón acelera el universo de forma casi exponencial. Al final de inflación las oscilaciones del campo alrededor del fondo del potencial recalienta el universo, generando toda la materia y energía que observamos. Llamamos Big Bang a este proceso y es tan violento que tuvo que producir enormes cantidades de ondas gravitacionales, que eventualmente serán detectadas.*

Bibliografía:
"El universo inflacionario", Alan Guth, Investigación y Ciencia, Nº 94, Julio 1984
"El universo inflacionario autoregenerante", Andrei Linde, Investigación y Ciencia, Nº 220, Julio 1995
"El universo elegante", Brian Green, Ed. Crítica, Drakontos (2006).

Juan García-Bellido Capdevila
Doctor Física Teórica
Catedrático, Universidad Autónoma de Madrid

Capítulo 60
¿Cómo se puede saber la edad del Universo? ¿Y su tamaño?
(Por Paola Marziani)

Esta es probablemente una de las preguntas más fascinantes que podemos plantear. Una primera respuesta ha llegado durante miles de años por el dogma, cerca de 6000 años, por la Biblia, que sin embargo marca el período de recolección proto-histórica de la humanidad. Otra edad se deriva por el estudio del radio-carbono en la tierra, que llega a 50.000 años. Isótopos con tiempos de decaimiento más largos permiten establecer la fecha de la roca más antigua de la Tierra en alrededor de 4 mil millones de años. Estas estimaciones son límites inferiores. Solo mediciones astronómicas pueden dar valores más realistas.

En primer lugar, una edad bien definida del Universo tiene sentido en el contexto de un universo en expansión (sería problemático si el universo fuera estacionario). Sin embargo, podemos aplicar la misma técnica de medición con isótopos radioactivos a las estrellas de la galaxia, ya que hay isótopos que tienen un tiempo de decaimiento que exceden los 40 mil millones de años. Para las estrellas más antiguas de la galaxia, pertenecientes a los cúmulos globulares del halo, se estima una edad de unos 14 mil millones de años. La teoría de la evolución estelar confirma ampliamente esta estimación. Pero, ¿y la edad del Universo en su conjunto?

Si extrapolamos la ley de Hubble descrita para objetos cercanos, suponiendo una constante de Hubble que no variara con la época cósmica, obtendríamos un valor menor de lo que se obtiene por otras técnicas de medición. Este resultado implica que la expansión del Universo no ha sido constante. Por lo que para conocer con precisión la edad del Universo, necesitamos saber la historia de su expansión. Actualmente, tenemos tres líneas de evidencia que sugieren un origen bien definido a partir de un estado denso y caliente:

1) el desplazamiento al rojo cosmológico

2) la radiación cósmica de fondo

3) y la nucleosíntesis de los elementos

La composición química del Universo primitivo era de hidrógeno, helio y pequeñas trazas de litio. La nucleosíntesis de elementos más pesados siguió de la generación de las primeras estrellas que se formaron con esta composición primordial, y la teoría de la evolución estelar define una escala temporal precisa. Sin embargo, hay otras limitaciones que surgen a partir de la necesidad de explicar la relativa homogeneidad e isotropía del Universo.

La teoría del Big Bang ha sido modificada por una fase especial de expansión superlumínica (es decir, que ocurre mucho más rápido la velocidad de la luz) para explicar la similitud de las regiones del espacio muy distantes. De este modo, el Universo disfrutó de una fase "tranquila," y la expansión se desaceleró frenada por la atracción gravitacional de la masa que el Universo mismo contiene.

Por debajo de un corrimiento al rojo z ~ 1 el Universo empezó a acelerar de nuevo. Esto fue deducido usando las Supernovas (cuya luminosidad intrínseca se conoce con gran precisión) que aparecieron más débiles de lo esperado (por lo que están más lejos) en el caso de una desaceleración continua.

La estimación actual de la edad del Universo incluye la información de la radiación cósmica de microondas, y las distancias de las supernovas y otro indicador de distancias para reconstruir la historia de la expansión del Universo y obtener un valor "de concordancia" de 13.8 mil millones de años (Gyr). Este valor está todavía sujeto a cambios porque nuevas campañas de observación y nuevos estudios están siendo ideados para reconstruir la historia de la expansión del Universo en detalle.

También se necesita la verificación de los supuestos y de los asuntos fundamentales sobre los que se basa la cosmología Λ-"cold dark matter": el origen de la energía asociada con la constante cosmológica que está impulsando la expansión en edades cósmicas recientes, así como su ecuación de estado.

La noción del tamaño del Universo no es obvia. Éste se expande a una velocidad que está cambiando con la época cósmica. Solo para mantenernos a nosotros mismos en un terreno firme, podemos considerar que ninguna señal puede viajar más rápido que la velocidad de la luz. Hemos estimado la edad del Universo como 13,8 Gyr. Sabemos que el tiempo que necesita la luz para viajar hacia nosotros desde el origen debe ser el mismo valor. Pero esto no es realmente el tamaño físico del Universo.

Durante la época inflacionaria se cree que, durante casi un instante justo después del Big Bang el Universo se puede haber expandido a un ritmo de expansión mucho mayor que la velocidad de la luz. Después de la fase inflacionaria la expansión del Universo se redujo a una tasa mucho más lenta; pero esta tasa estaba lejana de ser constante. La desaceleración y aceleración del Universo estaba gobernada por la fuerza de atracción de la gravedad y la fuerza de repulsión asociada con la llamada energía oscura.

Para tener una estimación razonable del tamaño del Universo debemos considerar lo que los astrofísicos llaman la distancia "comovil." Básicamente, es la distancia que se tendría si una cadena de observadores estuvieran de acuerdo para medir la distancia de las galaxias cercanas al mismo tiempo (claro que solamente es una idea conceptual que no puede realizarse en la práctica). En términos matemáticos, la expansión del Universo significa que las coordenadas están aumentando en un factor de escala que es una función del tiempo. La distancia comovil se calcula eliminando el efecto del factor de escala que depende del tiempo. La estimación de la distancia comovil claramente depende de la geometría del Universo, la cual es fijada por las fuerzas que gobiernan su expansión.

En realidad, su tamaño es de cerca 14.2 billones de parsec (Gpc), lo que corresponde a 46.3 mil millones de años luz, mucho más grande que el tiempo de viaje

de la luz. Se puede decir que el Universo es una esfera de este radio. Estamos en la condición de ver la luz emitida por las fuentes dentro de este ámbito, pero no más allá.

El radio de 14.2 Gpc puede ser considerado como un horizonte, más allá del cual hay una región no observable del Universo que no podemos conocer.

Se agradece la ayuda de las doctoras **Alenka Negrete** y **Alba Grieco** por una revisión precisa de este capítulo.

Bibliografía:
S. Weinberg, Gravitation and Cosmology, New York, J. Wiley & Son (1972)

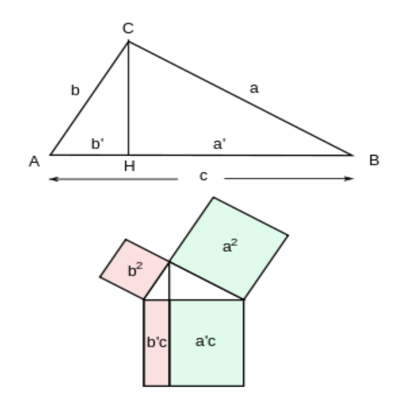
Paola Marziani
Doctora en Astrofísica
Investigadora / INAF - Osservatorio Astronomico di Padova

Capítulo 61
Filosofía: ¿Historia o historia inversa?
(Por Enrique Zuazua)

Si un extraterrestre nos visitara constataría que nuestra civilización ha evolucionado hasta una realidad admirable. La valoración que nuestro "estado del arte" le mereciera, dependería del nivel de evolución alcanzado por su propia civilización, pues todo es relativo, también en materia de progreso. Su percepción estaría condicionada por su capacidad de identificar la vida y de comunicarse en nuestras claves pues, no olvidemos, nuestra concepción de la vida y de la inteligencia está absolutamente condicionada por nuestra propia naturaleza. Ocurriría lo mismo si fuese alguno de nosotros quien visitase otro mundo.

La conocida novela "*El planeta de los simios*" (famosa también por la película de Charlton Heston (1968) y la posterior saga) del escritor francés Pierre Boulle (1912-2014) narra la situación inversa. El periodista humano Ulises Mérou, tripulante de la expedición a la estrella Betelgeuse en el año 2500 llega a Soror, uno de los planetas que orbita alrededor de la estrella, controlado por los simios y donde la raza humana vive en estado salvaje. Mérou debe demostrar a los simios que no es un animal, sino un ser inteligente y racional y se dirige así a la simia Doctora Zira:

¿Cómo no se me había ocurrido utilizar este medio tan sencillo? Tratando de recordar mis estudios escolares, tracé sobre el carnet la figura geométrica que ilustra el teorema de Pitágoras. No escogí este tema por casualidad. Recordé que, en mi juventud, había leído un libro sobre empresas del futuro en el que se decía que un sabio había empleado este procedimiento para entrar en contacto con inteligencias de otros mundos. [...]



Boulle enlaza así con la idea de que las Matemáticas constituyen el lenguaje del universo, tal y como estableció Galileo Galilei (1564-1642). Su necesidad emerge cuando el ser humano se plantea la posible comunicación con el más allá.

Si se tratase de un visitante extraterrestre que llegase a nuestro Planeta Tierra, su percepción cambiaría dependiendo del lugar de aterrizaje. No sería lo mismo hacerlo en un poblado rural del África subsahariana, donde todavía apenas llega el agua corriente o la luz, o en la azotea de uno de los rascacielos más modernos de Nueva York, Shanghái o Dubái. Y es que, en efecto, aunque hablamos con frecuencia de globalización, del estado del bienestar, de progreso, de civilización, al hacerlo, sin darnos cuenta, lo hacemos con referencia al estado social promedio que nuestro cerebro ha construido casi inadvertidamente, más con imágenes, percepciones sensoriales y estereotipos, por costumbre, que con datos cuantitativos contrastados y reflexiones críticas. De ahí que leer, pensar, estudiar, conversar, debatir, escribir, sea útil, con el objeto de aportar al cerebro todos esos detalles, con frecuencia contradictorios, que necesita para refinar su limitada percepción de la realidad y desarrollar la imprescindible capacidad de pensamiento crítico.

Vivimos en un mundo heterogéneo que nos vemos obligados a simplificar en exceso para poder visualizarlo en muchas ocasiones, como por ejemplo en el improbable contexto del encuentro del visitante venido de otro mundo. Pero, ¿cómo es el mundo en el que aterriza?, ¿con qué tipo de seres humanos se encuentra?, ¿cómo es la naturaleza en el entorno que pisa?

Es muy posible que al imaginar ese tipo de escena no solo estemos condicionados por nuestra propia percepción del entorno y de nuestra sociedad sino también por nuestras lecturas, históricas y fantásticas o, más posiblemente, por las imágenes que memorizamos de vivencias anteriores y también, cómo no, del cine. Sin ir más lejos, el famoso poema de Bernardo Atxaga *“37 Mugaz bestalde dudan lagun bakarrari”* (“37 Preguntas a mi único contacto al otro lado de la frontera”) recoge algunas de las paradójicas preguntas que podrían plantearse en ese encuentro, llenas de significado en sí mismas, sin respuesta: *“Mugaz bestaldean… ¿Arrain abisalek ba ahal dute aurresentipenik eguzkiaz…?”* (Al otro lado de la frontera… ¿Los peces abisales tienen presentimientos acerca del sol?) En otro conocido ejemplo, la película “ET” (Steven Spielberg, 1982) hubiese transcurrido de otra manera si, al inicio del guión, el extraterrestre llegado a nuestro planeta no se hubiese encontrado con un niño sensible, cariñoso, curioso y educado, sino con algún energúmeno que lo hubiese cazado para mostrarlo, inerte, como trofeo, en las redes sociales, ahora que ya no se pierde el tiempo en disecar las presas para exhibirlas en el majestuoso salón de casa.

La realidad que nos rodea no es única, en la medida que nuestra percepción de la misma depende de nuestra experiencia, de nuestro espíritu crítico y capacidad de análisis, y todo ello está íntimamente ligado a la educación recibida. Sea cual fuese el punto de llegada y origen de nuestro amigo del otro mundo, con independencia del clima imperante, del ruido, la contaminación y la compañía que le hubiese correspondido en su azaroso arribo, con certeza emprendería un proceso de análisis inverso, para intentar entender no solo el estado real de nuestra civilización sino el cómo hemos llegado hasta aquí, cuáles son sus fundamentos, su estructura, las reglas con las que nuestra gran tela de araña ha sido concebida en un laborioso proceso.

Poco a poco lanzaría un proceso de deconstrucción, con el objeto de entender todas las piezas del mecano que constituyen nuestra civilización actual y sus engranajes, nuestra organización social, la tecnología utilizada… De ese modo iría descubriendo la arquitectura de nuestras ciudades, y nuestras infraestructuras y, en un análisis regresivo, iría viendo cómo eran un siglo atrás, en blanco y negro, menudas, aún sin apenas

vehículos de motor, y mucho antes ciudades romanas o egipcias, y antes aún meros poblados de cabañas habitados por aquellos primeros homínidos que consiguieron salir de la caverna. Del análisis del mundo tecnológico que nos rodea, pronto descubriría que éste está basado en la informática y la robótica, nobles herederos de la Revolución Industrial y de las Matemáticas.

Cualquiera de esos caminos inversos recorridos le haría ir descubriendo el frondoso Árbol de la Ciencia, la Física, la Biología, la Química, las Matemáticas, hasta remontarse a aquéllos tiempos antiguos en los que todo el conocimiento se fundía en una única raíz plantada en el fértil tiesto de la insaciable necesidad humana de entender, de la sed del conocimiento, cuando se emprendió la gran aventura del pensamiento, basado en la creación de modelos cada vez menos contradictorios e imperfectos. Entendería también que nuestra sociedad reposa en gran medida en las Ciencias Sociales y las Humanidades, sobre las que hemos construido nuestro sistema económico, político y de justicia. Vería también que la moral, la ética, la política, el derecho y la religión se cruzan, intentando no solo dar estructura a nuestra sociedad, estableciendo límites, reglas y derechos, si no también intentando explicar racionalmente lo imposible: la propia existencia del universo y nuestro papel en él. Y en ese proceso de observación analítica identificaría la gran disciplina de la Filosofía, del latín *philosophĭa*, y este del griego antiguo φιλοσοφία, «amor por la sabiduría», como estudio de cuestiones fundamentales como la existencia, el conocimiento, la verdad, la moral, la belleza, la mente y el lenguaje. Ese proceso inverso de deconstrucción, de autopsia de nuestra sociedad, le haría ver también que, desde un perspectiva histórica, se arrancó de la nada para, poco a poco, a lo largo de los milenios, de siglos, llegar hasta donde estamos hoy, errando siempre, pero cada vez errando menos, mejor. Observaría también una sospechosa aceleración en nuestra capacidad de progresar, que corre el riesgo de llevarse por delante la sostenibilidad de nuestra especie sobre un Planeta cada vez más castigado. Constataría, necesariamente, que el nivel de progreso no es en absoluto homogéneo, no solo en términos de riqueza, tecnología, salud o, lo que coloquialmente denominamos "calidad de vida", sino también de organización social, de derechos, de libertades. No le resultaría difícil establecer conexiones entre el nivel de progreso más visible y el desarrollo intelectual de cada cultura, de cada país o región. Y pronto daría con una de las claves de lo que caracteriza y distingue a nuestra sociedad y que ordena a los países y las regiones en el ranking de las civilizaciones: su sistema educativo.

Llegados a este punto constataría que aquí estamos en una situación razonablemente buena pero alejada aún de los países líderes, de los que guían el destino del mundo y brillan más, ya sea en el ámbito de las nuevas tecnologías o en los Juegos Olímpicos. Comprobaría asimismo que en este ámbito de la educación las opiniones son muy diversas, que los consensos son casi imposibles, aunque se reconozca públicamente su necesidad, que coexisten subsistemas diversos incluso en sistemas de los que se podría esperar más cohesión, armonía y homogeneidad. Es lo que nos ocurre aquí, en Europa. De hecho le sorprendería que en las escuelas y en nuestro sistema educativo aún se debata sobre el papel que han de jugar materias tales como Religión, Ética o Filosofía, y que todas ellas sean mezcladas y relegadas a un cajón de asignaturas optativas, minoradas en el que el estudiante y/o sus padres han de elegir. Este hecho le resultaría especialmente paradójico pues, en este proceso de cuidadosa deconstrucción social, habría emergido de manera diáfana de entre los pilares de nuestra civilización, la disciplina de la Filosofía, la síntesis de los nobles esfuerzos polifacéticos humanos por dar un sentido al rol del ser humano en el universo, y estructurar su conveniencia en el

planeta, su fin, su destino. Cada materia puede ser contada de muchas maneras, en un orden cronológico o retrospectivo. Eso es sin duda debatible. Pero su necesidad no debería serlo.

Una de las grandes dificultades con la que nos encontramos los profesores es el vernos con frecuencia limitados por unos planes de estudio que nos empujan a presentar los contenidos desde el hoy hacia el ayer, a través del retrovisor, desnaturalizando así nuestras ciencias y su proceso de maduración, haciéndolas por tanto menos comprensibles y asimilables, al carecer de una motivación visible, palpable. Y es cada vez más unánimemente asumido que la perspectiva histórica es indispensable para estimular el interés del alumno y acentuar su capacidad de comprensión y análisis. Y un esfuerzo global por educar en lo que hoy sabemos necesariamente exige pasar por la Filosofía, lugar de encuentro tradicional de todas las ramas del conocimiento. Se puede debatir sobre cuáles han de ser los contenidos, el orden en que han de ser presentados, pero resulta paradójico que hoy tengamos que defender la necesidad de que en la escuela haya que reservar un espacio al pensamiento transversal y unificador que la Filosofía ha supuesto en nuestro desarrollo y devenir, su papel en la formación del espíritu crítico del ciudadano del mañana.

Como matemático he de recordar que la Filosofía y las Matemáticas, disciplinas hermanas, se encuentran en la lógica, la teoría del conocimiento y, en general, la abstracción. Más allá de las tradicionales disciplinas de contenido (ciencias físicas, ciencias de la vida, ciencias sociales) hay un conjunto de herramientas transversales que están presentes en cualquiera de esos dominios y en los que es imprescindible ser educado: las matemáticas, la informática, la filosofía y la estadística.

La formación y desarrollo del conocimiento requiere: a) acumular datos, b) estructurar información, c) construir el conocimiento, d) entender la realidad, e) alcanzar la sabiduría y madurez, y finalmente, f) transformar el mundo.

Y nuestro sistema educativo debería de estar orientado a ofrecer la mejor plataforma de lanzamiento en ese ambicioso plan que es, para cada uno de nosotros, un plan de vida. Así como las Matemáticas, la Informática y la Estadística son imprescindibles hoy en día, y así se reconoce en nuestro sistema educativo para entrenarnos y educarnos en el pensamiento deductivo, el procesamiento de la información, el pensamiento crítico y la inferencia basada en la probabilidad, la Filosofía ha pasado a ser la hermana pobre de la familia. Hoy, rodeados de móviles y cacharrería, el pensamiento crítico apenas tiene sitio pero es más necesario que nunca. El conocimiento no puede alcanzarse sin la sed por la sabiduría y, en ese camino que todos hemos de emprender, resulta indispensable ganar perspectiva, conocer los que ya fueron explorados, para elegir el más adecuado. Al final y al cabo el destino del humano es hacerse las preguntas adecuadas (Filosofía), planteando los problemas adecuados (Matemáticas). En la Antigua Grecia todas las ramas del saber se confundían. Y, hoy que la sabiduría está compartimentada, con frecuencia de manera artificiosa, es más necesario que nunca disponer de materias como la Filosofía, que ofrezcan una atalaya holística para contemplar todo lo que hemos construido, única manera de orientar adecuadamente los pasos futuros.

Es importante saber que podemos sobre-simplificar nuestro modelo del universo tanto como deseemos, poniendo como único referente el ser humano, o tal vez Dios o la Naturaleza, pero que una visión transformadora y actual pasa por la integración de todos

esos sub-modelos en uno más global. Y ese esfuerzo de fusión necesita de un taller, de un laboratorio específico que la Filosofía puede y debe representar.

Ser profesor de Filosofía es apasionante y a la vez sumamente difícil. Integrar todos los grandes hitos de la historia del pensamiento, desenmarañar el gran ovillo del conocimiento, es sin duda una tarea casi imposible. Pero no por eso deberíamos abandonar y renunciar a desempeñarla. Hoy, más que nunca, necesitamos de la Filosofía, ya sea contada en un orden cronológico, histórico, o en un sentido temporal inverso.

Hoy sabemos también que es imposible desvincular la Filosofía de la Ciencia, ambas en mayúsculas, en todas sus acepciones y versiones. Nos vemos abocados pues a retornar a las antiguas fuentes, a los orígenes, en que los grandes sabios no conocían fronteras entre las disciplinas del saber. Pero, en este afán, nos vemos condicionados por la falta de consenso, corriendo así el riesgo de que la beligerancia normativa de las distintas administraciones acabe arrinconando las materias que nos humanizan en el sentido más integrador del concepto.

No deja de ser paradójico que hoy que los más grandes sabios y científicos insisten cada vez más en la necesidad de unificar disciplinas para dar una coherencia global al conocimiento, el día a día nos arrastre en el sentido contrario. Volvamos, pues, a las fuentes para reconstruir nuestra historia y mirar al futuro con perspectiva y clarividencia.

Hay un verso de Joxean Artze en euskera que sintetiza a la perfección esa permanente necesidad: *"Iturri zaharretik edaten dut, ur berria edaten, beti berri den ura, betiko iturri zaharretik"* (Bebo de la vieja fuente, agua nueva, agua que siempre es nueva, de la vieja fuente de siempre).

Enrique Zuazua
Doctor en Matemáticas
Director de la Cátedra de Matemática Computacional, DeustoTech, Universidad de Deusto, Bilbao
Catedrático en Matemática Aplicada – Universidad Autónoma de Madrid

Capítulo 62
¿A qué se llama «radiación del cuerpo negro»?.
(Por José Ramón Martínez Saavedra)

Escribo estas líneas un caluroso día de junio en Madrid; todo aquel que haya pasado al menos un día de verano en esta ciudad sabe la tortura que supone salir a la calle pasadas las dos de la tarde: un asfalto tan caliente que se podrían incluso freír huevos en él, un aire tan irrespirable y pesado que se puede hasta cortar con un cuchillo y, además, un sol de justicia bajo el cual te achicharras vivo.

Y es que, digan lo que digan, no hay nada mejor que un día de verano en Madrid para acordarse de todos los procesos de transmisión de calor. El lector interesado puede observar, por ejemplo, la conducción de calor entre el asfalto y el aire que está en contacto con él; también puede estudiar la convección de este aire caliente, subiendo desde el asfalto, impactándole (sin piedad alguna) en la cara, haciéndole sudar. En ambos casos, el calor se transmite en un medio material: bien mediante el contacto entre los dos medios, bien mediante la aparición de corrientes dentro del material en sí.

Por último, no podemos olvidarnos del Sol, y de cómo esta inmensa bola de fuego es la responsable última de este calor (si se apagase, aunque fuese por un rato, a la hora de la siesta…). Sin embargo –a diferencia de los otros dos casos–, el calor del Sol no necesita de ningún tipo de medio material para transmitirse (y hacer reventar el termómetro), sino que calienta los objetos que tiene a su alrededor irradiándolos.

“Muy bien”, podrá decir el lector, “pero esta radiación, ¿de dónde sale? ¿Y qué tiene que ver con las transferencias de calor?”. Para ello, me temo que es necesario hacer un pequeño viaje al mundo subatómico, para tratar de explicar qué está ocurriendo ahí cuando hablamos de “transmisión de calor”: sin más dilación, reduzcamos nuestra escala en un factor de mil millones, a ver qué nos encontramos.

Una vez reducidos a esta escala, podemos fijarnos en los átomos del material: pequeñas bolas que… ¡se mueven! No demasiado, probablemente, pero parece que vibren alrededor de su posición de equilibrio: es esta energía asociada a la vibración de los átomos la que generalmente asociamos a escala macroscópica con la “temperatura” de un cuerpo. Con este modelo, que un cuerpo esté “más caliente” que otro quiere decir que sus átomos vibran más en comparación.

Esta imagen explica bastante bien los procesos de transmisión de calor por conducción: imaginad que ponemos en contacto dos cuerpos, uno muy caliente (átomos vibrando) y otro muy frío (átomos casi quietos). Al ponerlos en contacto, los átomos en la superficie empezarán a “colisionar” entre sí, intercambiando energía en el proceso: los átomos que antes estaban quietos empezarán a moverse (se pondrán calientes), mientras que los átomos que antes se agitaban furiosamente probablemente se agiten

bastante menos ahora (en otras palabras, se enfriarán): este proceso ocurrirá hasta que ambos materiales se agiten con la misma intensidad (es decir, que tengan la misma temperatura).

Para procesos de convección el mecanismo es bastante similar: la principal diferencia en este caso proviene del hecho de que estos procesos ocurren principalmente en fluidos, sustancias en las que los átomos tienen una mayor libertad de movimiento. En el caso de fluidos, los cambios en la temperatura suelen venir acompañados de cambios en la densidad, lo cual hace que se produzcan corrientes para redistribuir el fluido en orden de densidades. Estas corrientes mezclan regiones calientes (con átomos vibrantes) con regiones frías (con átomos casi estáticos), las cuales se equilibran entre sí posteriormente mediante procesos de conducción.

Por último, el modelo también es útil para explicar los procesos de emisión de radiación: cuando un átomo vibra no lo hace con velocidad constante, sino que sufre aceleraciones y deceleraciones constantemente: estas aceleraciones afectan a todas las partículas que conforman el átomo, las cuales tienen generalmente carga eléctrica (como el electrón y el protón). La teoría electromagnética predice –sin entrar en mucho detalle– que una partícula cargada que experimenta un movimiento acelerado emitirá radiación electromagnética; los parámetros de dicha radiación (su intensidad y su frecuencia) dependen del movimiento en sí: si la partícula vibra más, tanto la intensidad de la radiación (es decir, cuánto calor emite) como su frecuencia (es decir, en qué tipo de radiación electromagnética lo va a emitir) pueden cambiar.

La radiación de las partículas cargadas, además, se extiende por el espacio. Otra predicción —esta más intuitiva— de la teoría electromagnética es que las partículas cargadas pueden interactuar con los campos electromagnéticos que las rodean, intercambiando energía con ellos. Esto, aplicado a nuestro caso, quiere decir que la energía liberada por una partícula cargada (en forma de un campo electromagnético propagándose en el espacio) puede ser reabsorbida por otras, las cuales no tienen por qué estar en su entorno más cercano.

Por supuesto, también hay muchos otros procesos por los que un material puede emitir y absorber radiación, la mayoría de ellos asociados a transiciones electrónicas. Sin embargo, podemos intentar imaginarnos un sistema en el que eliminamos cualquier otra fuente de radiación. En este caso, la única fuente de radiación posible es la que hemos comentado anteriormente: la proveniente del hecho de que las partículas cargadas se agitan en el interior del material. Es a dicha radiación a la que nos referimos como "radiación de cuerpo negro".

Sin embargo, el origen del nombre "cuerpo negro" no proviene precisamente de considerar un emisor térmico, sino de otra definición que, a primera vista, no tiene nada que ver; como escribía Kirchoff en 1860 [1]:

La prueba que estoy a punto de ofrecer de la ley anterior asume que se puedan imaginar cuerpos tales que (para espesores despreciables) absorban completamente todos los rayos incidentes, y no reflejen ni transmitan ninguno de ellos. A dichos cuerpos los denominaré como "perfectamente negros" o, más brevemente, "cuerpos negros".

Esta definición, aunque en apariencia antagónica, resulta equivalente a la imagen que dimos anteriormente: un cuerpo que absorba toda la energía de la radiación, sin reflejar ni transmitir nada, implica que no hay procesos de radiación al margen de los térmicos. La energía de la radiación se convierte, en última instancia, en energía cinética de los constituyentes del cuerpo (si el material no sufre cambios en su estructura interna, por supuesto). El incremento en la energía cinética produce una agitación más intensa de las partículas, provocando una mayor cantidad de radiación térmica liberada al entorno. Dicha radiación tiene dos propiedades características: su dependencia exclusiva de la temperatura y su isotropía. La primera se refiere a que la radiación solo depende de la temperatura a la que esté el cuerpo negro; parámetros como su composición, su geometría, cualquier variable microscópica… son irrelevantes: si conocemos la temperatura del cuerpo negro, conocemos su patrón de emisión de radiación. Por otro lado, con "isotropía" nos referimos a que el tipo de radiación que emite el cuerpo es idéntica en todas las direcciones: da igual por dónde miremos, siempre encontraremos el mismo patrón de emisión tanto en intensidad como en la distribución en frecuencias de la luz.

Ambas propiedades se explican relativamente bien con nuestro modelo: si un material tiene la misma temperatura en todos lados, la energía cinética asociada a la agitación de las partículas tiene que ser la misma en prácticamente todos lados. Como dijimos anteriormente, cada temperatura tiene asociada una "energía cinética promedio" con la que se agitan los cuerpos: como el patrón de radiación depende únicamente de la cantidad de energía disponible de cara a que los componentes del material se agiten, el patrón de radiación debe depender de la temperatura única y exclusivamente. Por otra parte, este movimiento de agitación es aleatorio, y no tiene una dirección privilegiada; por tanto, en promedio se emitirá radiación en todas las direcciones posibles por igual, dando lugar al patrón isótropo de radiación que comentábamos anteriormente.

Otra propiedad extraordinariamente interesante de la radiación de cuerpo negro es su color: aunque parece de perogrullo preguntarse **"**¿de qué color es un cuerpo negro?", lo cierto es que el color de un cuerpo negro depende de la temperatura a la que esté: si aumentamos su temperatura, empezaremos a ver que el material pasa de negro a rojo, de rojo a amarillo, de amarillo a blanco y de blanco a… ¿azul? ¿Cómo que azul? La respuesta a esto, en el siguiente capítulo.

Notas:
[1] «The proof I am about to give of the law above stated, rests on the supposition that bodies can be imagined which, for infinitely small thicknesses, completely absorbs all incident rays, and neither reflect nor transmit any. I shall call such bodies *perfectly black*, or, more briefly, *black* bodies». Extraído de "On the relation between the radiating and absorbing powers of different bodies for light and heat. The London, Edinburgh and Dublin philosophical magazine and journal of science (Taylor & Francis) 20 (130)"

José Ramón Martínez Saavedra
Magíster en Fotónica
Doctorando — ICFO-Institut de Ciències Fotòniques

Capítulo 63
Al calentar una sustancia se pone primero roja, luego naranja, después amarilla, pero a continuación blanca. ¿Por qué no sigue el espectro y se pone verde?

(Por José Ramón Martínez Saavedra)

En el capítulo anterior discutíamos sobre las posibles formas de transmisión de calor —conducción, convección y radiación—, y tratamos de explicarlas recurriendo a un modelo microscópico de los materiales. Este modelo, aunque simple, era capaz de justificar (con más o menos gracia) por qué los cuerpos emitían radiación electromagnética por el simple hecho de estar a una cierta temperatura. Discutimos también qué se entiende por un "cuerpo negro", así como las propiedades más importantes de la radiación térmica generada por esta clase de cuerpos: su isotropía, o la capacidad de emitir idénticamente en todas las direcciones del espacio; y su dependencia exclusiva de la temperatura, o el hecho de que el modo de emisión de radiación depende únicamente de la temperatura a la que esté el cuerpo, y de ningún parámetro más.

Sin embargo, cerré el capítulo con una pregunta en el aire que, a simple vista, parece absurda: "¿de qué color es un cuerpo negro?". Aunque parece del mismo estilo que el típico acertijo "¿de qué color es el caballo blanco de Santiago?", deja de serlo en el momento en que probamos a calentar una sustancia demasiado; haciendo esto vemos que, acompañando al aumento de la temperatura, va asociado un cambio en el color del material. En castellano decimos que se pone "al rojo vivo", y no solo metafóricamente: la sustancia adquiere un tono rojizo, tanto más brillante (y amarillento) cuanto más caliente está la sustancia en sí.

Dicho esto, podría tener sentido hablar del color de un "cuerpo negro" después de todo. Sin embargo, para poder hablar con propiedad, sería necesario decir antes qué se entiende por "color", desde un punto de vista físico… y enlazar "color" con alguna de las propiedades de una onda electromagnética no es precisamente intuitivo. Para solventar este problema, podemos estudiar un ejemplo distinto, y tratar de trasladarlo a las ondas electromagnéticas: el caso de las ondas sonoras.

Las ondas sonoras tienen una serie de propiedades comunes con las ondas electromagnéticas: ambas se describen mediante su amplitud (de alguna forma proporcional a la intensidad de la onda), y ambas tienen una frecuencia a la que dicha amplitud oscila. En el caso de las ondas sonoras, dicha frecuencia de vibración está íntimamente ligada con el *tono* con el que percibimos el sonido; es decir, cómo de grave o de agudo es. Los tonos más graves están asociados a frecuencias bajas, y se hacen más agudos a medida que aumentamos la frecuencia de la vibración. Eventualmente, la frecuencia de la vibración será tan alta que nuestros oídos dejarán de ser sensibles a dicha vibración: en otras palabras, "no oímos" los sonidos más agudos que un cierto tono.

Algo similar ocurre cuando tenemos ondas electromagnéticas: en este caso, nuestros ojos actúan como nuestros sensores. La información sobre el "tono" de la onda se codifica en forma de colores, siguiendo la escala cromática del arcoíris: las ondas electromagnéticas de menor frecuencia se verán rojas; a medida que aumentemos la frecuencia irán pasando por el naranja, amarillo, verde, azul y añil, hasta llegar al violeta para las frecuencias más altas. Igual que en el caso acústico, el ojo no es capaz de detectar todo tipo de radiación electromagnética, sino que es sensible únicamente en una región (generosa, eso sí), del espectro electromagnético: cualquier onda con frecuencias fuera de este rango no será detectada por el ojo, y un cuerpo que solamente emita fuera de la ventana de sensibilidad del ojo se verá de color negro.

Por supuesto, esta visión es una versión simplificada de la denominada "teoría del color": saber de qué color se verá un material en función de su espectro de radiación es una de las principales áreas dentro del campo de la óptica. En general, la radiación visible emitida por un cuerpo no es monocroma (incluso aunque lo veamos de un solo color), sino que está compuesta de múltiples colores, cada uno de ellos presente en un porcentaje variable: la combinación —y posterior interpretación por parte del cerebro— es la que, en última instancia, produce el color en sí.

Este es el caso, por ejemplo, de los cuerpos negros: estos no emiten su radiación en una sola frecuencia, sino en todas las frecuencias a la vez (decimos que presentan un "espectro continuo" de radiación). Sin embargo, la intensidad relativa a dicha frecuencia depende de forma directa de la temperatura a la que se encuentre el cuerpo en cuestión. Si representamos en una gráfica la intensidad de la radiación de cuerpo negro para cada frecuencia, encontraremos gráficas similares a las representadas en la figura siguiente

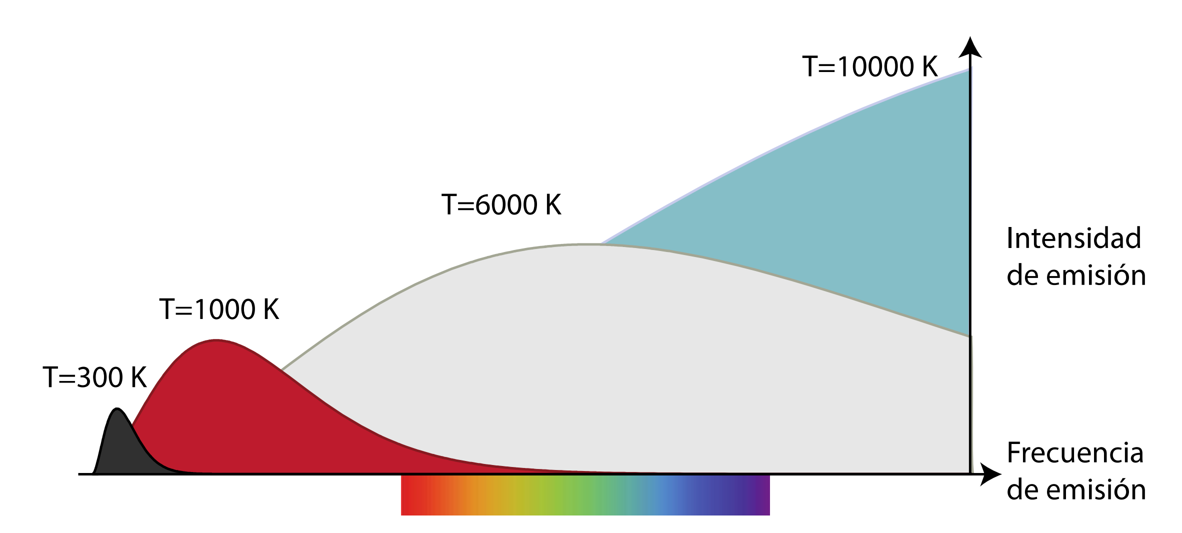


Figura 1 *Espectro de radiación de cuerpo negro en función de la frecuencia. La región del espectro visible al ojo humano aparece a modo de guía. El color de cada una de las curvas es el color aproximado de un cuerpo negro a esa temperatura.*

Como primer detalle importante, podemos ver como el máximo de emisión en frecuencias se desplaza hacia frecuencias más altas a medida que la temperatura sube; del mismo modo, la cantidad de energía liberada al entorno aumenta también con la temperatura [1].

Otro detalle interesante es la "cola" de la distribución; es fácil ver que dicha cola se extiende hacia frecuencias más altas, y su extensión depende de la temperatura. Así,

para una temperatura de 300 K (es decir, unos 30 ºC, o temperatura ambiente) toda la radiación se encuentra contenida en una región no accesible para el ojo humano: por ello, no somos capaces de detectar radiación proveniente de estos objetos, y los vemos de color negro. No obstante, si seguimos aumentando la temperatura hasta 1000 K (unos 730 ºC, bastante típica de materiales en hornos de fundición), vemos como parte de la energía emitida se introduce en la región visible, por la región de baja frecuencia: como resultado, comprobaremos que el material comienza a brillar con un característico color rojizo.

Si seguimos aumentando la temperatura hasta casi 6000 K (una temperatura típica, por ejemplo, en la superficie del Sol), tendremos el máximo de la distribución justo a la mitad del espectro visible: como resultado, tenemos que todos los colores del espectro visible son emitidos más o menos con la misma intensidad; su combinación, como podéis imaginar, genera un color blanco-amarillento, siendo el blanco la interpretación por parte de nuestro cerebro de una combinación de colores en la que todo el espectro es excitado por igual.

Por último, y como colofón al capítulo, cabe preguntarse qué ocurre si seguimos calentando el material (asumiendo que podemos, por supuesto). En este caso, el máximo de radiación se desplazará hacia las regiones de rayos ultravioleta del espectro, y el balance de colores en el visible se desplazará hacia el azul; sin embargo, a diferencia de para temperaturas muy bajas, en ningún caso dejará de haber radiación en la región visible: cuanto más calentemos el material, más azul y más brillante se hará.

Visto todo lo anterior, solo me queda daros un consejo: si veis un material brillando, es mejor no tocarlo: brille rojo, amarillo, blanco o azul, seguramente va a estar demasiado caliente para cogerlo con las manos.

Notas:
[1] Estas son las denominadas “ley de Wien” y “ley de Stefan-Boltzmann”, respectivamente; ambas se pueden derivar fácilmente de la expresión analítica para las curvas de radiación,

$$P(\nu, T) = \frac{2h\nu^3/c^2}{\exp\left(h\nu/k_B T\right) - 1},$$

conocida como “ley de Planck”.

José Ramón Martínez Saavedra
Magíster en Fotónica
Doctorando — ICFO-Institut de Ciències Fotòniques

Capítulo 64
¿Qué son las líneas del espectro de un elemento? ¿Tiene alguna utilidad su estudio?
(Por Pablo López Tarifa)

Toda materia que conocemos interactúa con la luz, ya sea reflectando, absorbiendo o emitiéndola. El estudio de estas interacciones es lo que se conoce como espectroscopia, y sus técnicas son fundamentales en los laboratorios de todo el mundo. Por ejemplo, la espectroscopia ayuda a los científicos a saber de qué clase de átomos está compuesto un material recién sintetizado, cuántas partículas de contaminantes hay en una determinada muestra, o de qué están compuestas las atmósferas de otros planetas. También los médicos se sirven de técnicas espectroscópicas, como los rayos-X y pruebas NMR, para hacer más precisos sus diagnósticos. Nosotros, en nuestra vida cotidiana, también usamos controles remotos, microondas, teléfonos móviles o conexiones wi-fi, que son aparatos diseñados para emitir o recibir luz de un determinado "tipo".

Para entender esto último necesitamos dar una definición más formal de lo que comúnmente conocemos como luz. En su sentido físico, la luz se puede definir como la forma que tiene la energía de propagarse en el espacio. Según la teoría ondulatoria, la luz es una onda electromagnética, es decir, es la oscilación de un campo eléctrico y otro magnético perpendiculares entre sí (y perpendiculares a su vez a la dirección de desplazamiento). Al igual que sucede con las ondas que podemos generar en el agua de una piscina, las ondas electromagnéticas también tienen crestas y valles. A la distancia que separa dos crestas se le conoce como longitud de onda, llamada λ, magnitud que se usa para clasificar las distintas ondas del espectro electromagnético. De esta forma, en el extremo más alejado nos encontramos las ondas gravitacionales, con longitudes de onda entorno al radio terráqueo. Estas ondas son particularmente débiles ya que, como establece la teoría corpuscular de la luz, la cantidad de energía transportada por la onda/fotón es inversamente proporcional a su longitud de onda. Si seguimos disminuyendo la distancia entre las crestas, pasamos a las ondas de radio, con longitudes de onda en torno a kilómetros; seguidas las ondas de microondas, con longitudes de onda de centímetros. Acortando la longitud de onda nos encontramos en la franja de infrarrojo con longitudes en un rango de milímetros a micro metros (10^{-6} m). Seguidamente hay una franja muy pequeña, entre 700 y 400 nano-metros (10^{-9} m) donde se encuentran las longitudes de onda que nuestro ojo es capaz de percibir. A este rango se le denomina espectro de visible y muchos de los objetos que nos rodean absorben en él, y por ello decimos que tienen color. El aparato usado para detectar qué longitudes de onda de visible absorbe un material en particular, es el llamado espectroscopio de visible. El espectro continua con longitudes de onda más pequeñas, ya no visibles, agrupadas en diversas franjas: ultravioleta, rayos X,…

Vista desde el nivel atómico, cuando las ondas electromagnéticas interactúan con un objeto, éstas lo hacen con sus electrones. Como sabemos gracias a los postulados

de la mecánica cuántica, los electrones tienen energías cuantificadas, esto es, solo pueden tener ciertos valores discretos. Así, si la frecuencia de la onda electromagnética, ν, (que se relaciona con la longitud de onda y la velocidad de la luz según $\nu=c/\lambda$) es equivalente a la energía que separa dos niveles electrónicos del material, el fotón desaparece (se absorbe), promocionando sus electrones a un nivel excitado. Si la frecuencia no cumple este requerimiento, la onda electromagnética pasa a través del material sin ser absorbida. De esta manera, comparando las longitudes de onda de la luz que irradiamos sobre un material y el remanente una vez ésta ha pasado a su través, veremos que faltan las longitudes de onda de los fotones que han sido absorbidos. Si descomponemos esta luz transmitida en función de sus longitudes de onda y las imprimimos en una placa fotográfica, obtendremos lo que se denomina espectro de absorción. Fijándonos con detenimiento, nos encontraremos con líneas negras (ausencia de luz) en aquellas longitudes de onda absorbidas. Como hemos visto, el número y posición de estas líneas depende de los niveles electrónicos del material, y éstos a su vez de la composición atómica de éste. Así el espectro de una sustancia es único, al igual que lo son nuestras huellas dactilares. Si un elemento o compuesto en particular está presente en una muestra dejará su huella en el espectro, esto es, absorberá determinadas longitudes de onda que lo harán reconocible.

Este mismo razonamiento se puede extender a otro tipo de espectroscopias como la de infrarrojo o microondas, muy utilizadas en el estudio de los astros. Así por ejemplo, descomponiendo la luz que nos llega a la Tierra de los diferentes planetas del sistema Solar, podemos saber si, por ejemplo, el metano, CH_4, está presente en sus atmósferas: Lo único que tenemos que hacer es buscar dos bandas intensas a 1300 y 3000 cm^{-1} de energía [1].

Hoy en día, muchos grupos de investigación trabajan en el desarrollo de nuevas técnicas espectroscopias con el objetivo de revelar, a un nivel mucho más fundamental, cómo la radiación electromagnética interactúa con los electrones y átomos que la componen. Por ejemplo, usando pulsos ultra-cortos en el rango de los atto-segundos (10^{-18} s), los científicos han sido capaces de "filmar" el movimiento electrónico en el rango de los femto segundos (10^{-15} s) [2]. Este estudio ha probado que las técnicas desarrolladas son capaces de "manipular" los electrones de pequeñas moléculas orgánicas, induciendo localizaciones de carga en partes concretas. El sueño de controlar las reacciones químicas, desde su nivel fundamental, en el rango de los atto-segundos, donde tiene lugar el movimiento electrónico que crea y rompe enlaces, parece más cerca que nunca.

Bibliografía:
[1] NIST Chemistry WebBook: http://webbook.nist.gov/
[2] F. Calegari et al. *Science* "Ultrafast electron dynamics in phenylalanine initiated by attosecond pulses" Vol. **346**, 6207, 336-339.

Pablo López Tarifa
Doctor en Química
Investigador Postdoctoral en la Vrije Universiteit, Amsterdam

Capítulo 65
¿Qué es la luz polarizada?
(Por José Manuel Llorens Montolio)

La naturaleza de la luz ha sido un problema muy controvertido en física. Históricamente se ha atribuido a la luz propiedades características, tanto de ondas como de partículas. Isaac Newton consideró que los fenómenos de propagación rectilínea, reflexión y refracción en una superficie que separa dos medios podían explicarse suponiendo que la luz estaba compuesta de numerosas partículas. Estos fenómenos se observan cuando se enciende una linterna y vemos un claro círculo iluminado en una pared situada frente a nosotros. El mero hecho de poder ver ese círculo es indicativo de la capacidad de reflexión de la luz, pues ésta sigue una trayectoria que se inicia en la linterna, rebota en la pared y llega a nuestros ojos. El fenómeno de refracción se observaría al considerar una pared transparente, sea por ejemplo, un acuario lleno de agua.

Sin embargo, la luz también experimenta fenómenos de interferencia más propios de las ondas. Aunque también son muy comunes, su observación resulta más sutil. A modo de ejemplo tenemos las franjas claras y oscuras ondulantes que aparecen al plegar una cortina traslúcida situada frente a una ventana iluminada. La correcta descripción de todos estos fenómenos llegó de la mano de James Clark Maxwell, quien estableció cuatro ecuaciones que aún hoy se siguen empleando sin mayores modificaciones a su formulación original y son la base de la óptica. Dichas ecuaciones describen un carácter puramente ondulatorio de la luz.

Isaac Asimov se pregunta sobre la naturaleza de la luz polarizada. La cuestión está bien fundamentada, pues en nuestra vida cotidiana asistimos a fenómenos ondulatorios más evidentes que la luz, como son las olas del mar o la vibración de una cuerda de guitarra y, sin embargo, no es común oír hablar de la polarización. A pesar de ello, la polarización es una propiedad intrínseca de una onda y nos indica la dirección en la que se está produciendo la oscilación. Es por ello que Isaac Asimov recurre al símil de una cuerda atada a un árbol para ejemplificar la dirección de vibración o polarización de la onda. Así, si hiciésemos oscilar nuestro brazo de arriba a abajo veríamos que la cuerda vibra en una dirección paralela al árbol. Podríamos decir que la onda está polarizada paralelamente al árbol, es decir, verticalmente. Si lo hiciésemos de izquierda a derecha, hablaríamos de una onda polarizada perpendicularmente al árbol, esto es, horizontalmente. Finalmente, si moviésemos nuestro brazo en cualquier dirección podríamos hablar de una onda no polarizada con respecto al árbol. Sobre cómo controlar la polarización, Isaac Asimov también nos da una solución. Haciendo pasar la cuerda por un objeto que bloquee todas las direcciones de vibración menos una conseguiremos seleccionar la dirección de polarización de la vibración de nuestra cuerda. Para ello, la propuesta de Asimov es utilizar una valla y hacer pasar la cuerda por el hueco entre los listones verticales. Podemos considerar la valla como un “polarizador” de nuestra

cuerda. De forma análoga, aunque mediante un mecanismo físicamente diferente, la disposición de los átomos en un cristal puede hacer que solo una dirección de vibración de la luz pueda propagarse libremente a través de él.

Las aplicaciones que tiene el control de la polarización son innumerables y cualquier aficionado a la fotografía conoce las virtudes de colocar un polarizador en su objetivo (eliminar reflejos indeseados e intensificar los colores de las superficies). Sin embargo, en la actualidad, y gracias a la gran difusión de los teléfonos móviles y pantallas planas, el control de la polarización está más presente en nuestro día a día que en el pasado. En estos dispositivos se pueden distinguir diferentes capas. En la primera se encuentra una fuente de luz blanca. Sobre ésta se sitúa un primer polarizador lineal, que tiene un efecto similar al de la valla de la que hemos hablado antes. Este primer polarizador obligará a que la luz que lo atraviese esté polarizada en una dirección. Posteriormente está colocado un cristal líquido. Aquí es donde aparece el elemento activo de control de la polarización. Unas moléculas muy alargadas se orientan en una determinada dirección con un elemento eléctrico (generalmente un transistor) que nuestra pantalla controla a voluntad. Dependiendo de si se aplica o no un voltaje, estas moléculas se orientarán o no en una dirección. En el estado de apagado, las moléculas son capaces de rotar la dirección de polarización. En el estado de encendido, las moléculas no actúan sobre la dirección de polarización. Un segundo polarizador será el encargado de no dejar pasar la luz en el estado apagado y de permitirlo en el estado encendido.

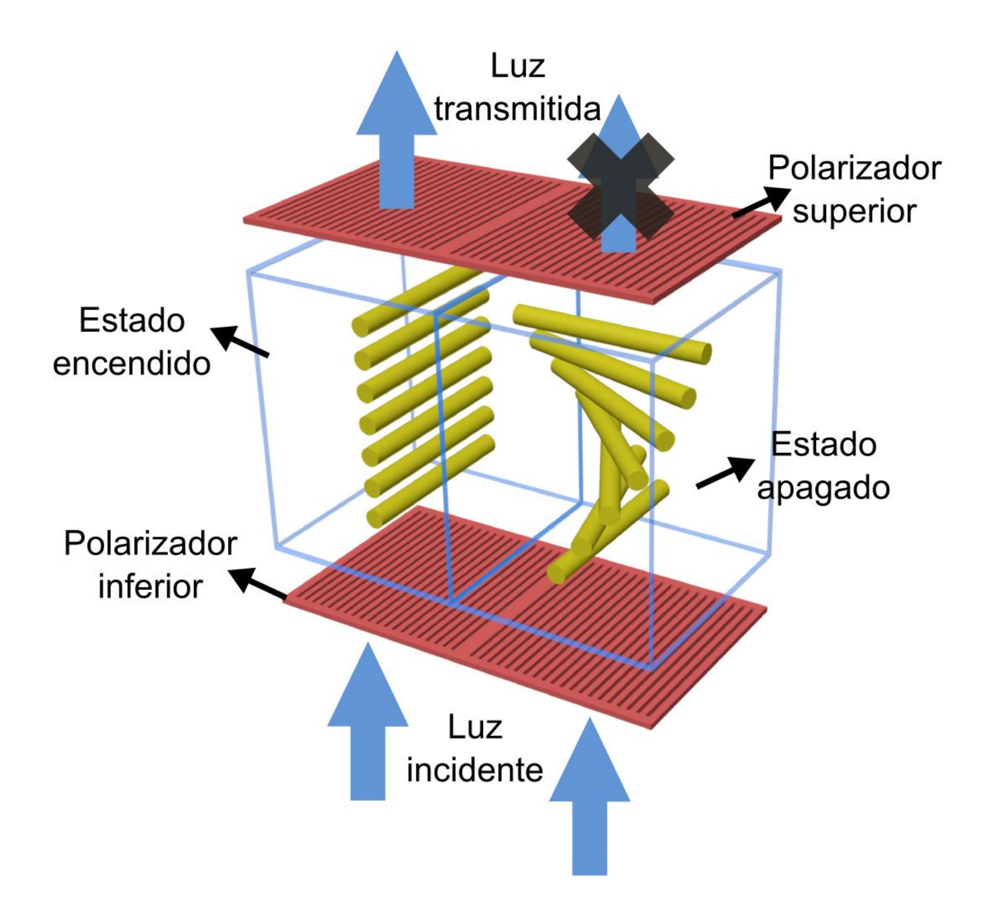


Figura 1. *Esquema del control de la polarización realizado en un dispositivo con una matriz de cristal líquido. La luz incide desde abajo y dependiendo del estado encendido o apagado la luz se transmitirá hacia el observador.*

Para facilitar la comprensión de este mecanismo, hemos incluido un esquema de funcionamiento en la Figura 1. Ésta muestra la posición de los polarizadores (en rojo) y la orientación de las moléculas en el cristal líquido (cilindros amarillos). La celda a derecha se corresponde con el estado apagado y la celda izquierda con el estado encendido. Finalmente, unos filtros de color rojo, verde y azul serán los encargados de seleccionar el color que acabará siendo transmitido, formándose una imagen en color en nuestros ojos. La implementación puede ser diferente a la que hemos descrito, sin embargo la esencia es la misma: utilizar la polarización para bloquear o permitir el paso de luz. El buen control que se haga de esa polarización será lo que proporcione una pantalla totalmente negra en el estado apagado en lugar de un tono gris oscuro. De igual forma, al inclinar la pantalla de nuestro móvil muchas veces observamos fenómenos curiosos, como inversión de colores, reducción de gama o, simplemente, no vemos nada. Una de las batallas de los gigantes tecnológicos consiste en controlar la polarización para fabricar pantallas con las mejores prestaciones. En un futuro podremos disponer de pantallas donde el apagado y encendido se efectúe directamente sobre el elemento emisor de luz y se pueda eliminar el control de la polarización del dispositivo. Aunque conceptualmente pueda parecer más sencillo, su implementación aún requiere solventar muchos problemas para reducir su coste.

La polarización de la luz también juega un papel importante en dispositivos en los que, en lugar de controlar la emisión de luz, se busca su absorción. Éste es el caso de las células solares. En toda célula solar existe una región en la que queremos que toda la luz que le llega del Sol sea completamente absorbida. Al absorber la luz en el material, su energía es recogida por los electrones que pueden así generar trabajo. La luz del Sol llega sin una dirección de vibración definida, esto es, es luz no polarizada. Cuando esta luz incide sobre la célula solar, una parte es reflejada hacia el aire, otra parte es absorbida por la célula y, dependiendo del espesor que tenga, habrá otra parte de luz que se absorba en la estructura que la soporta. Exceptuando la luz absorbida en la célula, el resto son pérdidas de energía. Hay por tanto una cantidad de luz que nuestra célula no puede absorber. Para aumentar la cantidad de luz absorbida podemos hacer uso de la naturaleza ondulatoria de la luz y colocar un filtro a la entrada que obligue a la luz a no ser reflejada y que además desvíe su dirección de propagación, de forma que tenga que recorrer un camino más largo antes de encontrarse con la cara posterior de la célula. Al aumentar su recorrido dentro del material absorbente de la célula, llegará menos luz a la cara posterior.

El diseño de estos filtros recibe el nombre de técnicas de atrapamiento de luz. La disposición de patrones ordenados en la cara iluminada de la célula potencia esos fenómenos de interferencia que estamos buscando. Sin embargo, al igual que ocurría con la valla y la cuerda, nuestro filtro puede que ayude a la propagación de la luz polarizada en una dirección pero resulte un impedimento para la luz que oscila en otras direcciones. Actualmente se trabaja en el diseño de filtros que presenten cierto desorden, de forma que el beneficio se obtenga para cualquier dirección de vibración y no solo para unas pocas. De nuevo, el control de la polarización resulta clave para el rendimiento de un dispositivo.

Desde un punto de vista fundamental, el análisis de las soluciones de las ecuaciones de Maxwell no deja de sorprender a los científicos. Además de la polarización, ha sido posible identificar en la luz propiedades solo conocidas hasta ahora en partículas como, por ejemplo, el momento angular y el espín. Más aún, ha sido posible identificar un efecto relativista como la interacción spin-órbita en luz

propagándose en materiales. Estos fenómenos son objeto de estudio actualmente y abren la puerta a nuevas preguntas y aplicaciones que, cuándo Isaac Asimov escribió su libro, pertenecían más a la ficción que a la ciencia.

José Manuel Llorens Montolio
Doctor en Física
Científico Contratado
Instituto de Microelectrónica de Madrid (IMM-CSIC)

Capítulo 66
La luz roja es la menos desviada al pasar por un prisma, pero la que más se desvía al pasar por una red de difracción. ¿Por qué esa diferencia?

(Por José Manuel Llorens Montolio)

En este reto planteado por Isaac Asimov, vamos a retomar el problema de la naturaleza de la luz que planteamos en el capítulo anterior. En primer lugar tenemos que entender por qué la luz de color rojo se desvía menos que la de cualquier otro color. Este efecto se popularizó enormemente al ilustrar la banda Pink Floyd la portada del álbum "The Dark Side of the Moon" con un haz de luz blanca incidiendo en una de las caras de un prisma y obteniendo en la opuesta un arco iris. El dibujo de la portada muestra como cada uno de los colores se propaga en una dirección diferente, siendo el haz de la luz de color rojo el menos desviado respecto a la dirección de propagación del haz incidente. Este efecto de cambio de dirección de propagación recibe el nombre de refracción y sucede cuando la luz entra en un medio en el que su velocidad de propagación es diferente de la del medio en el que se estaba propagando antes. Isaac Asimov nos explica con un símil marcial el fenómeno de la refracción de la luz. Antes de continuar vamos a incidir en este punto, puesto que luego describiremos unos nuevos materiales cuyas propiedades son una generalización del fenómeno de refracción. Apoyándonos en los gráficos mostrados en la Figura 1, tenemos que la luz se propaga uniformemente por un medio rápido, que hemos ilustrado con el fondo blanco. Cuando un rayo de luz alcanza la intercara con un medio en el que se propaga más lentamente (medio azul) éste se ve forzado a desviarse, pues de otra forma se perdería la uniformidad de la propagación de los rayos.

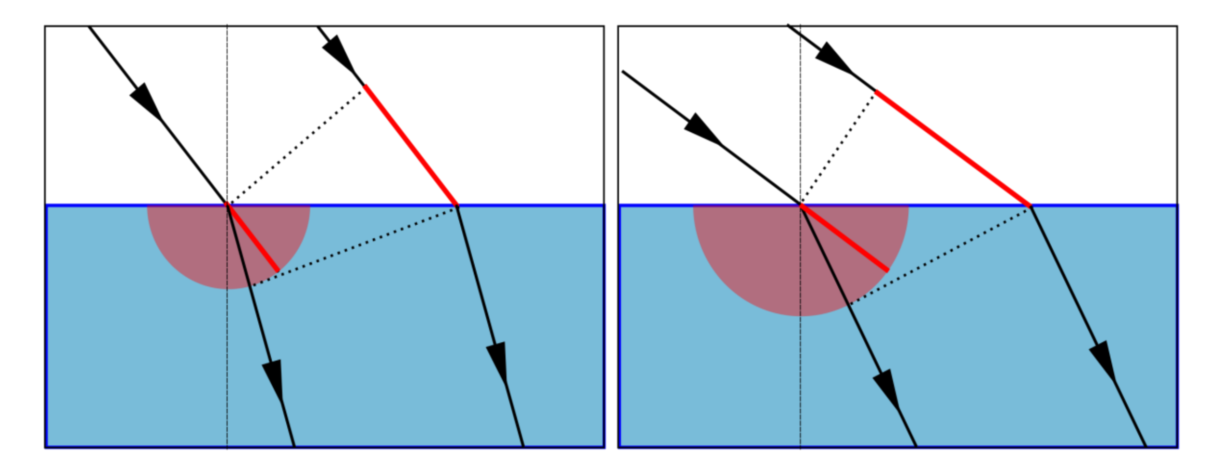


Figura 1. *Descripción geométrica de la refracción de la luz al pasar de un medio rápido (fondo blanco) a uno lento (fondo azul). Se ha ilustrado para dos ángulos de incidencia diferente.*

La uniformidad está representada por la línea de puntos. Suponiendo que sean soldados en marcha esta línea marcaría el frente de la fila de soldados. Con una línea

roja hemos marcado la distancia que recorren los soldados en los extremos de la fila. Se puede apreciar como el soldado que sigue en el medio rápido recorre una distancia mayor que el soldado que ya ha entrado en el medio lento, sea por ejemplo un lodazal. La semicircunferencia roja indica los diferentes caminos de igual longitud que podría haber recorrido el primer soldado que entró en el lodazal. Como no está permitido perder la uniformidad de la fila, el soldado tomará aquella dirección que le una a su compañero en el extremo opuesto a través de la línea perpendicular. Hemos ilustrado este fenómeno de refracción con dos ángulos de entrada en el material lento diferentes.

Desviándonos un poco de la pregunta original de Isaac Asimov, vamos a presentar unos materiales que son capaces de desviar la luz de forma que cambie no solo su dirección sino también su orientación. Hasta ahora hemos hablado de haces que se propagan de arriba hacia abajo y de izquierda a derecha. Este comportamiento es el que siempre vemos en la naturaleza y el que nos resulta más conocido. Sin embargo, recientemente se han producido en el laboratorio una combinación de pequeñas estructuras embebidas en una matriz cuya respuesta ante una iluminación es diferente a la de cualquier material natural. Estos materiales se conocen como metamateriales. Aunque son muchas sus sorprendentes propiedades, aquí introduciremos los materiales de índice de refracción negativo. El índice de refracción es el cociente entre la velocidad de la luz en el vacío y la que ésta tiene en un material. En los gráficos mostrados más arriba el índice de refracción, que resulta ser el cociente entre la longitud de las líneas rojas, es 1,75. La luz en el medio azul es 1,75 veces más lenta que en el blanco. ¿Qué sentido tiene hablar de luz propagándose en un medio con índice de refracción negativo? Uno estaría tentado de decir que ahora la luz se propagaría en sentido contrario. Este efecto no sería muy novedoso, ya que estaríamos hablando de la luz reflejada en la intercara entre ambos materiales. Lo que nos dice este índice negativo es que la luz se propaga en el medio como si se propagase en sentido contrario. Para aclarar este punto introduciremos un nuevo gráfico en la Figura 2.

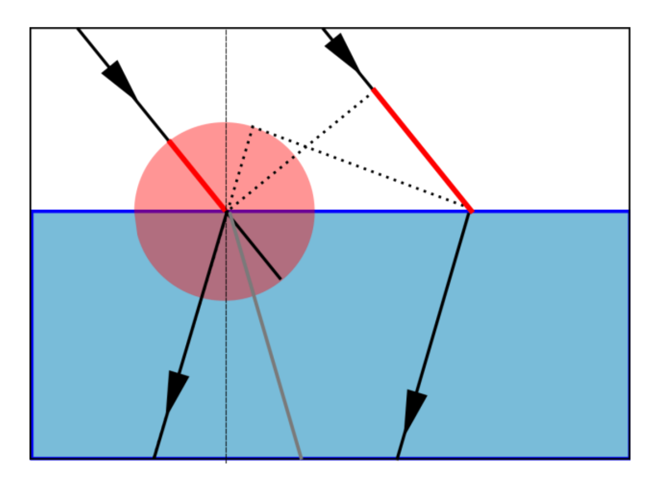


Figura 2. *Descripción geométrica de la refracción de la luz al pasar de un medio rápido con índice de refracción positivo (fondo blanco) a uno lento con índice de refracción negativo (fondo azul).*

Partimos de una construcción análoga al caso de un índice positivo. Dibujamos la línea roja más corta en el medio rápido y buscamos la dirección que debe tomar de forma que no se pierda la uniformidad de la columna. La línea gris indica el camino que hubiese tomado para un índice positivo y la negra para el medio de índice negativo. Como vemos, la columna ahora cambia su orientación, al entrar en el medio azul se

propaga de derecha a izquierda. Ahora nuestro soldado se mueve en el lodazal como si viniese de una columna de soldados que avanzase en una dirección de igual inclinación pero desde el lado derecho. En otras palabras, de una columna de soldados que fuese el reflejo en un espejo colocado perpendicular a la linde de lodazal (línea vertical en el gráfico). Este simple hecho abre nuevas vías de control de la propagación de la luz. Por ejemplo, es posible diseñar una lente plana capaz de enfocar un punto luminoso situado a un lado de la lente en otro punto situado en el lado opuesto. Con un material convencional no sería posible, pues en ningún momento se cambia la orientación de la propagación y la luz emitida por un punto se dispersa sin volver a ser enfocada de nuevo.

Volviendo nuevamente a la pregunta planteada por Isaac Asimov, la luz roja seguiría, en principio, siendo la menos desviada en el prisma, tanto para un índice positivo como negativo. Como se explica en el texto original, es la diferencia de longitud de onda y su interacción con los electrones de los átomos del material la responsable de que sea más o menos desviada. Así pues, si la longitud de onda es muy grande frente al tamaño típico de los átomos (caso de la luz roja) la interacción con éstos será muy pequeña y el índice de refracción será menor que para la longitud de onda más corta (caso de la luz azul). Cuando hablamos de difracción la situación es la opuesta. Recordemos que, haciendo una gran simplificación, podemos hablar de la difracción como la capacidad de la luz para rodear un obstáculo que encuentra a su paso. En el caso de una partícula, ésta experimenta una colisión en un determinado punto del perímetro del obstáculo y se desviará con una trayectoria rectilínea. Para el caso de una onda, la interacción es un poco más difícil de visualizar. Ésta viene determinada, en gran medida, por el tamaño relativo entre la sección transversal del objeto y la longitud de onda de la luz. Resulta entonces que esa relación de tamaños es el parámetro responsable de que el haz se desvíe más o menos. Ahora la luz se comporta como una onda y cuánto más pequeña sea su longitud de onda frente al objeto, menos se desviará. Encontramos entonces que la luz azul se desvía menos que la luz roja por su reducida longitud de onda. Sin embargo, al hablar de metamateriales tenemos un material con objetos de mucho mayor tamaño que el tamaño del átomo aunque no lo suficientemente grandes como para que sean capaces de difractar la luz. Es por ello que en este régimen de tamaños intermedios es arriesgado hacer generalizaciones y serán los experimentos y la resolución de las ecuaciones de Maxwell lo que nos indique cuánto y por qué la luz roja o azul se desvía más o menos.

José Manuel Llorens Montolio
Doctor en Física
Científico Contratado
Instituto de Microelectrónica de Madrid (IMM-CSIC)

Capítulo 67
¿Qué pasa con la energía cuando dos haces luminosos interfieren de manera destructiva?
(Por Olga Caballero Calero)

Aunque quizá no seamos conscientes de ello, todos hemos visto o estudiado alguna vez el fenómeno de la interferencia destructiva: quizá las ondas que produce una piedra al lanzarse a un lago, quizá el experimento de la doble rendija, o al sacudir una cuerda (o al tocar un instrumento de cuerda)… En nuestra vida diaria, somos conscientes de que necesitamos el plato giratorio del microondas para conseguir que la comida se caliente homogéneamente, aunque quizá no nos hayamos preguntado porqué. Esto ocurre porque dentro del horno microondas, las ondas están en modo estacionario, esto es, la suma de la onda emitida y la onda reflejada dentro de la cavidad forman una onda estacionaria con máximos de intensidad (donde la comida se calienta mucho) y nodos o mínimos (donde no se calienta nada), y por ello tenemos que girar el plato para homogeneizar la temperatura. Pero, ¿realmente no hay energía en los nodos? ¿Dónde se va la energía?

Como bien dice el título, ésta es la pregunta que ha motivado este capítulo, es decir, saber qué ocurre con la energía cuando dos haces de luz interfieren entre si y producen oscuridad. De hecho, como veremos más adelante, no podemos tener el caso de dos haces de luz que interfieran y produzcan total oscuridad, esto es, no existe la interferencia destructiva total. De hecho, lo que ocurre es que la interferencia de dos haces coherentes produce oscuridad local, y si miramos el sistema global, no solamente la zona de interferencia, veremos que la energía no desaparece, si no que se "reorganiza".

Un ejemplo clásico de este efecto de oscuridad local producido por interferencia de haces de luz es el experimento de la doble rendija: una fuente de luz se coloca a una cierta distancia de una pantalla con dos rendijas. Si se coloca una segunda pantalla al otro lado de la fuente de luz, se observará el patrón de interferencia creado por la doble rendija. Este patrón se compone de franjas luminosas y oscuras, que corresponden a interferencia constructiva y destructiva, respectivamente. Lo que ocurre en este patrón es un caso claro de redistribución de la energía, es decir, que la suma total de intensidad del patrón iguala a la suma de la intensidad de luz que pasa por ambas rendijas. De hecho, las franjas son el doble de intensas que en el caso de tener la luz proveniente de una sola rendija. En conclusión, la energía se conserva, pero se reorganiza espacialmente.

El otro caso de interferencia, esto es, aquel que produzca interferencia destructiva en todas partes, no es posible. Intuitivamente, podríamos pensar en un experimento sencillo, con una simple cuerda, en el que pudiéramos generar una interferencia destructiva. Imaginemos que tenemos una cuerda que se agita de tal modo que se crea una onda harmónica que se propaga a lo largo de la misma, con lo que

transportaría una cierta energía. Entonces, podemos imaginarnos otra onda de la misma frecuencia y amplitud pero en desfase con la anterior. Esta segunda onda también transportaría energía. Si sumamos ambas ondas, resultaría que la cuerda dejaría de moverse, sin frecuencia ni amplitud alguna y, por lo tanto, ¡haciendo desaparecer la energía! ¡Interferencia destructiva total! Por supuesto, hay un pequeño problema con este razonamiento, y es que en el ejemplo no estamos teniendo en cuenta por ninguna parte cómo se han podido generar estas dos ondas en la misma cuerda. Tras la generación de la primera onda, en algún punto de la cuerda, deberemos aplicar un movimiento tal que se genere esta segunda onda. Una manera sería hacer un movimiento que cancelara la primera onda, es decir, haciendo una fuerza que trabaje en contra de la primera onda. El resultado sería que, a partir del punto en el que se está haciendo esta fuerza, la cuerda no se movería. Sin embargo, la actuación de esta fuerza actuaría reflejando la onda incidente, por lo que no añade energía al sistema total. La energía que "desaparece" se ha transformado en trabajo para cancelar la primera onda.

Sería fácil intentar imaginar un experimento similar con luz, colocando dos láseres perfectamente alineados que emitieran en direcciones opuestas. La solución más sencilla a este problema es que si consiguiéramos interferencia destructiva en todas partes, los láseres anteriormente mencionados no podrían emitir luz, ya que su modo de emisión sería un modo prohibido y el fenómeno físico que produce la emisión de estos láseres sería reemplazado por otra forma de energía, tal como emisión de fonones, es decir, calentamiento del láser, por ejemplo. Esto puede resultar difícil de imaginar en un primer momento, pero para cualquiera que haya trabajado con láseres es fácil recordar qué ocurre cuando se tiene un sistema perfectamente alineado, lo que supone que alguna reflexión del sistema pueda entrar de nuevo en la cavidad, desestabilizando el láser y haciendo que la propia emisión del láser se comporte de modo extraño. Esto se produce porque los láseres tienen una arquitectura muy estudiada en cuanto al diseño de su cavidad, de modo que ésta responde a un cierto medio amplificador. La luz no se crea automáticamente, si no que hay una serie de fenómenos de acumulación y emisión que han de producirse de una manera determinada. Es decir, tal y como yo lo comprendo, no hay ninguna posibilidad de crear una interferencia totalmente destructiva con luz.

De todos modos, alguien que tenga algo de experiencia con interferómetros, por ejemplo, podría tirar por tierra toda mi interpretación de redistribución espacial de la energía. Por ejemplo, en el caso de un interferómetro de Michelson (ver la figura 1), el patrón de interferencia producido aparece como una diana. Cuando uno de los brazos del interferómetro se bloquea, vemos toda el área de la diana homogéneamente iluminado. Si tomamos la intensidad total con interferencia y lo comparamos con la suma de la intensidad que proviene de los dos brazos del interferómetro, no se corresponden. De hecho, ¡hemos perdido energía! ¡Esto sí podría ser un caso de destrucción de energía! Pero (siempre hay un pero para preservar la energía), hemos dicho anteriormente que la energía se reorganiza "globalmente" (he reorganizado la frase también). ¿Qué significa "globalmente"? Significa que tenemos que mirar por todas partes para buscar la energía perdida antes de afirmar que ha desaparecido. En este caso, tenemos un efecto similar al de la cuerda que hemos discutido anteriormente. En este tipo de interferómetros, hay otro patrón de interferencia que no solemos tener en cuenta y que se refleja en el patrón de interferencia. Si se combina este patrón (también llamado antisimétrico, ya que corresponde a la imagen invertida de la diana que observamos) con el patrón de interferencia original, la intensidad resultante sí corresponde a la suma de las intensidades de ambos brazos del interferograma. Ya hemos encontrado la energía perdida.



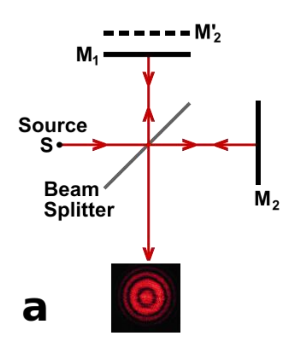


Figura 1: *Interferómetro de Michelson (fuente: Wikipedia), S es la fuente,* M_1 *y* M_2 *son los espejos y* $M`_2$ *es la imagen de* M_2*.*

Por todo esto, la respuesta a la pregunta propuesta en el título es que, cuando dos haces de luz interfieren destructivamente para crear oscuridad, la energía… se va a otro lado. Esto no es más que otro modo de decir que la ley de la conservación de la energía siempre se cumple, como es de esperar. Cuando la energía desaparece en algún punto, se puede encontrar siempre en otro lugar, quizá como otro tipo de energía. Para concluir, y cambiando un poco las palabras de Einstein, "Durante el resto de mi vida reflexionaré sobre *dónde* está la luz".

Olga Caballero Calero
Doctora en Ciencias Físicas
Investigador Contratado
Instituto de Microelectrónica de Madrid (IMM-CSIC)

Capítulo 68
¿Qué son la nanociencia y la nanotecnología? ¿Cómo afectarán a nuestras vidas?

(Por José Miguel García Martín)

En griego antiguo, el término “νανοζ” significaba “enano”. En ciencia, desde mediados del siglo XX, el prefijo “nano” se emplea para designar la milmillonésima parte de una unidad, de modo que el nanómetro equivale a 0,000000001 metros. Para el lector profano, esta sucesión de ceros terminada en un uno puede que signifique poco, así que trataremos de dar cuenta con unos ejemplos de lo pequeña que es esa distancia. Podemos empezar cogiendo una regla decimal, como esa que llevan los niños y las niñas a la escuela: la separación entre dos rayitas pequeñas es un milímetro (1 mm), y si fuésemos capaces de dividir ese espacio en un millón de partes iguales, tendríamos un nanómetro (1 nm). Otra forma de obtenerlo sería arrancándonos un pelo de la cabeza: el grosor de ese pelo es del orden de una décima de milímetro (0,1 mm), así que bastaría trocearlo en cien mil partes iguales para tener 1 nm… ¿Por qué esa ínfima distancia ha cobrado tanta importancia en nuestros días, donde “nanociencia” y “nanotecnología” son términos que leemos en los periódicos o escuchamos en los telediarios con cierta frecuencia? Pues por un doble motivo:

1) Porque es adecuada para describir el tamaño de los constituyentes elementales de la materia y de la vida. El radio de un átomo es del orden de décimas de nm, por lo que el tamaño de las moléculas es del orden del nm: la de glucosa, por ejemplo, mide 1 nm de largo. Los ribosomas también miden 1 nm aproximadamente, el diámetro de la doble hélice del ADN es de unos 2 nm, un anticuerpo mide unos 10 nm, un virus una centena de nm y las bacterias son de varios centenares de nm.

2) Porque los avances en instrumentación permiten ahora visualizar y manipular esas entidades. Los microscopios electrónicos (desarrollados a partir de 1930) y los microscopios de sonda de barrido (surgidos en los años 80) permiten caracterizar los nanoobjetos [1]. Algunos de estos últimos, como el microscopio de efecto túnel y el de fuerzas atómicas, pueden emplearse para coger átomos y moléculas de una superficie y trasladarlos de un sitio a otro: se habla entonces de nanomanipulación. Además, gracias a varias técnicas de nanolitografía (óptica, electrónica, por iones…) y a diversos métodos de fabricación (síntesis química, electroquímica, sol-gel, epitaxia de haces moleculares, ablación láser, pulverización catódica…), es posible fabricar una gran variedad de nanomateriales.

Siendo rigurosos, deberíamos diferenciar entre los términos mencionados, nanociencia y nanotecnología, en función de que estemos aludiendo a investigación básica o aplicada. Así, la nanociencia puede definirse como el estudio de los fenómenos fundamentales a escala atómica, molecular y macromolecular, donde las propiedades difieren significativamente de aquellas que se manifiestan a escala macroscópica. Por su parte, la nanotecnología hace referencia al diseño, la fabricación, la caracterización y la

aplicación de estructuras, dispositivos y sistemas mediante el control de la forma y el tamaño a escala nanométrica [2]. El interés por ambas disciplinas se debe a que permiten obtener nuevos materiales y nuevos dispositivos que manifiestan propiedades novedosas o que consiguen realizar nuevas funciones. Sin embargo, la pregunta básica que el lector puede estarse planteando es: ¿por qué? Es decir: ¿por qué los objetos o sistemas de escala nanométrica se comportan de forma distinta a como lo hacen los del mundo macroscópico al que estamos acostumbrados?

En primer lugar, se debe a la importancia que adquieren las superficies de esos objetos o sistemas. Los átomos de la superficie están enlazados de modo distinto a como lo hacen los del interior (de un modo naif, podemos decir que están sujetos de diversa forma y por ello son más susceptibles a cambios) y además están en contacto con el entorno. Muchas propiedades dependen de cuál es la proporción de átomos que están en la superficie, como por ejemplo la reactividad química con el medio que los rodea. Imaginemos un cubo de Rubik en el nanomundo, donde cada uno de los 27 cubitos de colores que lo constituyen fuese un átomo. Pues bien: de esos 27 átomos, 26 de ellos estarían en la superficie, y solo 1, el del centro del cubo, estaría en su interior. Podemos afirmar por tanto que la relación superficie/volumen es de 26/27=0.96. Imaginemos ahora un dado macroscópico, con aristas de 1 cm en las que en cada una de ellas hay colocados 10 millones de átomos (asumo para el ejemplo que cada átomo ocupa un nanómetro cúbico de volumen). El lector puede comprobar que ese dado está formado por 10^{21} átomos (un uno seguido de veintiún ceros), de los cuales "solo" $6x10^{14}$ (un seis seguido de catorce ceros) están en la superficie. Por tanto, en ese dado la relación superficie/volumen es de apenas 0,0000006. ¿Cuál será entonces más reactivo, el nanocubo de Rubik o el dado de un centímetro cúbico? O dicho de otro modo: si el lector fuese la persona responsable de compras de una empresa que necesita un determinado catalizador para sus procesos (un catalizador es una sustancia que acelera una reacción química estando en contacto con sus reactivos pero sin participar en ella), ¿cómo querría que se lo suministrasen, en forma de nanocubos o en forma de dados?

Fenómenos sorprendentes asociados a efectos superficiales en la nanoescala podemos encontrarlos en la naturaleza. La salamanquesa puede caminar por paredes y techos porque las terminaciones de sus patas contienen fibras de tamaño nanométrico muy juntas entre sí a modo de tapiz: cada fibra interacciona de forma atractiva (técnicamente: por fuerzas de Van der Waals) con la superficie por la que se desplaza, y al ser tan alto el número de fibras existentes, el efecto global consigue vencer a la gravedad. En los laboratorios se desea replicar este comportamiento, es decir, se intenta sintetizar materiales "bioinspirados" en esas patas que contengan nanofibras y que presenten súper-adhesión: lo que Spiderman consigue en los cómics o en las películas puede que no esté tan lejos…

Pero además, el pequeño tamaño de los nano-objetos ofrece otras ventajas. En componentes electrónicos, menor tamaño implica que los portadores de carga eléctrica (electrones o huecos) deben recorrer menores distancias, lo que se traduce no solo en más rapidez sino también en mayor eficiencia, pues la probabilidad de que en su camino se encuentren con un defecto es lógicamente menor. Y por supuesto, al disminuir el tamaño cobran importancia los fenómenos cuánticos, y aunque su explicación se escapa al alcance de este capítulo, cabe señalar que algunas de las aplicaciones que ya usamos en nuestra vida diaria hacen uso de ellos. Por ejemplo, las cabezas lectoras de los discos duros de ordenador actuales emplean el efecto de la magneto-resistencia túnel: este fenómeno se da en multicapas de unos pocos átomos de espesor en forma de sándwich,

donde “el pan” son electrodos magnéticos y “el relleno” una lámina no metálica, cuya resistencia eléctrica depende del estado magnético relativo entre esos electrodos (imanaciones paralelas o antiparalelas).

Este campo, el de las tecnologías de la información y las comunicaciones, es el primero donde los avances en nanociencia han conseguido ya plasmarse en diversas aplicaciones [3]. Las memorias magnéticas de nuestros ordenadores tienen cada vez mayor capacidad, esto es, permiten almacenar más información en menos espacio físico. Para ello, se logran fabricar nuevos materiales que contienen nanoimanes más pequeños en los que la información guardada en código binario en forma de estados magnéticos (polo norte-polo sur para los “unos”, polo sur-polo norte para los “ceros”) se mantiene estable. La incesante miniaturización de los transistores es la que posibilita que tengamos memorias flash o USB que podemos llevar cómodamente en nuestros bolsillos. Los avances en fotónica permiten incrementar cada año la capacidad de almacenamiento óptico de datos y la velocidad de las telecomunicaciones por fibra óptica. Y a más largo plazo, se espera que la computación cuántica multiplique exponencialmente las capacidades de cálculo actuales.

También son relevantes las aplicaciones de la nanotecnología en el campo de la energía, para lograr un mayor aprovechamiento y un menor impacto medioambiental. Por ejemplo, en lo que atañe a la conversión de luz en energía eléctrica, han aparecido en los últimos años dos nuevas familias fotovoltaicas, las células solares sensibilizadas con colorante y las células basadas en perovskitas, que son alternativas muy prometedoras frente a los paneles solares de silicio convencionales, y donde la nanoestructuración de los materiales se traduce en una mejora de la eficiencia. Otro caso significativo de esto son los nanomateriales termoeléctricos, que permiten transformar eficazmente en electricidad el calor residual de diversos procesos o dispositivos, como el de los tubos de escape de los vehículos con motor de combustión. Es importante señalar que diversos gases contaminantes, como el dióxido de carbono, pueden capturarse y almacenarse empleando nanomateriales porosos. Y además, se investiga en procesos que logren la fotosíntesis artificial, es decir, la conversión potencial de ese dióxido de carbono a metanol u otros productos químicos de valor añadido, empleando para ello la luz del sol y un fotocatalizador semiconductor nanoestructurado.

Asimismo, la nanotecnología está ofreciendo soluciones en otras áreas de gran interés social como son las ciencias de la vida y la biotecnología. Ya existen en el mercado cremas de protección solar que emplean nanopartículas de dióxido de titanio para filtrar la radiación ultravioleta. Y en el caso de implantes ortopédicos, se están analizando diversos recubrimientos nanoestructurados y biocompatibles con propiedades antibacterianas, que paliarían notablemente los problemas de infección y rechazo. Cuando además se emplean estrategias de “funcionalización”, esto es, de dotar a las nanoestructuras de nuevas funciones complementándolas con otras entidades, las posibilidades se amplían. En la lucha contra el cáncer, hay planteadas diversas rutas basadas en nanopartículas que se funcionalizan para que se adhieran a las células cancerígenas, por ejemplo acoplándoles un anticuerpo que se une específicamente a una proteína que solo existe en dichas células. Mediante la aplicación de un agente externo, como puede ser luz infrarroja en el caso de nanopartículas plasmónicas (de metal noble) o un campo magnético alterno si las nanopartículas son magnéticas, se puede conseguir un calentamiento muy localizado que provoque la apoptosis o muerte celular programada: son los denominados tratamientos por hipertermia. También se investiga en el suministro dirigido de fármacos, en el que las nanopartículas hacen de portadoras

de medicamentos (es decir: se funcionalizan previamente para llevar fármacos adheridos) que se administran localmente, en la zona del organismo que es necesario curar, mediante la activación de un determinado mecanismo: un cambio en el pH, un calentamiento, etc.

Por último, conviene señalar que también se investiga acerca de los posibles riesgos derivados de la fabricación y el uso de los nanomateriales. Aunque éstos se fabrican en su gran mayoría empleando sustancias autorizadas por la normativa vigente, en principio inocuas para los seres vivos y el medioambiente, es cierto que su pequeño tamaño puede dotarles de alguna propiedad perjudicial que no posean en la escala macroscópica. Por ello, se trabaja en nanoseguridad en cuatro temáticas principales: la clasificación de los nanomateriales (por su morfología, por su composición química, por su complejidad y funcionalidad, por su afinidad a entidades biológicas), la exposición y transformación de los mismos (cómo se liberan de los productos de que forman parte y cómo se modifican a lo largo de su ciclo de vida, incluyendo su posible reciclaje), toxicología y ecotoxicología (a qué organismos pueden afectar, a partir de qué umbrales, qué alteraciones pueden producir, cómo se eliminan del organismo), y por último la predicción de riesgos (identificación y estimación de los riesgos, implementación de herramientas de control, confección de una base de datos común sobre nanomateriales y riesgos asociados). En la Unión Europea, diversos grupos de trabajo que se ocupan de estos temas han creado el Cluter en Nanoseguridad, que se presenta como un foro donde discutir e informar al público general de modo responsable [4].

En definitiva, tras esta breve panorámica de aplicaciones en diversos campos con gran interés social y económico como son las tecnologías de la información, el aprovechamiento energético y las ciencias de la vida, confío en que el lector quede convencido de la importancia que los materiales y dispositivos en la nanoescala tienen actualmente, que se incrementará de forma notable en años venideros.

Notas:
[1] El Premio Nobel de 1986 reconoció estos descubrimientos: la mitad fue para Ernst Ruska, inventor del microscopio electrónico, y la otra fue compartida entre Gerd Binnig y Heinrich Rohrer, inventores del microscopio de efecto túnel.
[2] Definiciones adaptadas de la Royal Society (http://www.nanotec.org.uk/).
[3] De hecho, cuando en 2007 el francés Albert Fert y el alemán Peter Grünberg recibieron el Premio Nobel por haber descubierto la magneto-resistencia gigante, antecedente de la magneto-resistencia túnel, la Real Academia Sueca de Ciencias afirmó que este descubrimiento era "una de las primeras aplicaciones reales del prometedor campo de la nanotecnología".
[4] http://www.nanosafetycluster.eu/

Bibliografía:
"Nanosciences: the invisible revolution", Christian Joachim y Laurence Plévert, World Scientific (2009).
"La nanotecnología", Pedro Serena, Los Libros de la Catarata (2010).
"Nanosafety in Europe 2015-2025: Towards Safe and Sustainable Nanomaterials and Nanotechnology Innovations", elaborado por diversos científicos del Finnish Institute of Occupational Health y disponible en: http://www.ttl.fi/en/publications/Electronic_publications/Nanosafety_in_europe_2015-2025/Documents/nanosafety_2015-2025.pdf

José Miguel García Martín
Doctor en Ciencias Físicas
Científico Titular del CSIC
Instituto de Microelectrónica de Madrid (IMM-CSIC)

Capítulo 69
¿Qué es el grafeno y por qué nos importa?
(Por María A. H. Vozmediano)

69.1 ¿Qué es exactamente?

El grafeno es un cristal compuesto por una sola capa plana de átomos de carbono que se colocan en los vértices de una red de panal de abeja, como las mallas de los gallineros tradicionales.

¿Qué tiene de especial?: ¡Una sola capa! Es decir, es un material de espesor atómico. Para hacernos una idea de lo que esto significa, digamos que una hoja de papel normal o un cabello humano tienen aproximadamente un millón de capas atómicas. Ningún material puede ser más fino.

69.2 ¿Cómo se obtiene?

La síntesis que dio lugar al premio Nobel de física 2010 se realizó por exfoliación a partir del grafito. Esta forma del carbono está compuesta de capas monoatómicas unidas entre sí muy débilmente. Usando el ahora famoso método "del papel cello", Kostya Novoselov y André Geim de la Universidad de Manchester fueron capaces de aislar una sola capa de grafito. También consiguieron identificarla, extraerla del adhesivo, depositarla sobre un sustrato, poner contactos y demostrar inequívocamente que habían obtenido grafeno.

El tamaño típico lineal de las muestras empezó siendo de micrometros (milésimas de milímetro) y alcanza el milímetro en la actualidad. Si bien es cierto que cualquier humilde laboratorio puede sintetizar grafeno (también lo hay en los trazos de la escritura del lápiz y en la comida carbonizada que se pega a la sartén), su manipulación requiere un nivel alto de experiencia experimental en el ámbito de la nanotecnología (un nanometro es una milésima de micra).

Hoy en día se obtiene grafeno de mayores dimensiones con métodos más convencionales. Mediante la deposición química de grafito vaporizado (los átomos de carbono evaporados calentando grafito a 1.000 grados centígrados se van depositando espontánemente sobre ciertos metales como el cobre, en una red hexagonal) se obtienen metros cuadrados pero con muchos defectos y a veces más de una capa. Por sublimación del silicio en el carburo de silicio se obtienen muestras de cientos de micras.

69.3 ¿Por qué tanta emoción?

Tres aspectos:

69.3.1. Conceptual.

a) Se pensaba que no podían exisitir cristales bidimensionales. Aparte de algunos teoremas que niegan la existencia de dichos cristales (en condiciones ideales), la experiencia era que a la naturaleza no le gustan las dos dimensiones. La construcción sistemática de hidrocarburos planos añadiendo átomo a átomo al cristal siempre da lugar, a partir de un número crítico de átomos, a una reorganización espontánea de los átomos para llenar un espacio tridimensional. La exfoliación de materiales hechos de capas como el grafito también fallaba al bajar de un número crítico de capas. ¿Cómo se las apaña el grafeno no solo para existir sino para ser muy estable e inmune a ataques ambientales químicos o térmicos? La clave está en la naturaleza de los enlaces covalentes que unen entre sí los átomos de carbono, mucho más fuertes que la energía térmica que desestabiliza otros compuestos bidimensionales. La geometría especial de la red de panal de abeja es también clave para la mayoría de las propiedades del material.

b) Comportamiento relativista de los portadores de carga. Las partículas elementales que constituyen los materiales viven en cautividad. Sus comportamientos difieren, en general de sus compañeras libres. En particular, los electrones responsables de la conductividad eléctrica en los metales clásicos (aluminio, cobre) se ven casi atrapados en la jaula de los iones lo que dificulta sus movimientos y la relación entre la energía y el momento es diferente de la de sus compañeros libres. También su masa "efectiva" es mayor. En resumen, se mueven siguiendo las leyes de la mecánica cuántica no relativista. Recordemos que la mecánica relativista (clásica o cuántica) se aplica a las partículas que van a velocidades próximas a la de la luz. La sorpresa del grafeno es que sus electrones, aunque se mueven casi tan despacio como los de los otros metales, (1/300 veces la velocidad de la luz), por la estructura de la jaula en que viven, han perdido su masa y por tanto no se pueden describir con la mecánica clásica. Lo que es más sorprendente es que también desafían el paradigma de la mecánica relativista. Las partículas sin masa se han de mover a la velocidad de la luz. Este comportamiento fascinante bajo el punto de vista científico, está también en el origen de muchas de las propiedades físicas especiales.

69.3.2. Propiedades físicas inesperadas.

El grafeno, debido a su espesor y a las peculiaridades de la red de panal de abeja, tiene una serie de comportamientos denominados "superlativos" por sus descubridores. Además de generar una auténtica revolución en la física del estado sólido, estas propiedades son la clave para generar aplicaciones. Las más relevantes son:

a) Electrónicas: la calidad de un conductor se mide a menudo en términos de algo llamado "movilidad". También es importante la intensidad de corriente que puede soportar un conductor - sin quemarse-. La movilidad depende mucho de la temperatura. El grafeno compite en ligera inferioridad con los mejores componentes electrónicos - semiconductores clásicos como el arseniuro de galio - a muy bajas temperaturas pero es imbatible a temperatura ambiente. Esto hace en particular que se haya podido observar por primera vez un fenómeno de física cuántica (efecto Hall cuántico) a temperatura

ambiente. También se observa cómo los electrones de grafeno son capaces de "atravesar barreras" de manera cuántica (paradoja de Klein).

b) Morfológicas: aunque en los primeros tiempos del grafeno llamaron la atención sus propiedades electrónicas, pronto se vio que las propiedades mecánicas del material son posiblemente más espectaculares. En particular en este ámbito se encuentran los siguientes "superlativos":

- Fuerte: aguanta la tracción sin romperse. Es uno de los materiales más resistentes conocido, 200 veces más que el acero. Es también más duro que el diamante.
- Elástico: se puede deformar hasta un 20% volviendo a la situación original. En comparación, el acero se puede deformar aproximadamente un 0.25%.
- Impermeable: ni siquiera el helio es capaz de atraverar una membrana de grafeno.

c) De especial interés es la particular interrrelación que se da en el grafeno entre las dos propiedades anteriores: las deformaciones elásticas afectan a las propiedades electrónicas de manera muy particular: se acoplan a los electrones igual que los fotones (campos gauge). De nuevo aquí aparece una física propia de las partículas elementales: las deformaciones de la red actúan de manera parecida a un campo magnético ficticio. Esta propiedad permite detectar deformaciones de la red con experimentos de difracción de electrones.

69.3.3. Aplicaciones.

Muchas de las aplicaciones más o menos inmediatas del grafeno están heredadas del grafito que era ya uno de los materiales más útiles conocidos -junto con la fibra de carbono o el carbono amorfo-. Algunas de las aplicaciones más prometedoras ya en curso utilizan el grafeno en combinación con tecnologías estándar para mejorar las propiedades de otros materiales. En particular es un excelente aislante térmico, muy buen lubricante, y se puede usar para dar rigidez a otras estructuras sin aumentar el peso o el tamaño. Sus propiedades electrónicas lo hacen útil en células fotovoltaicas, baterías, antenas. Las propiedades fotoeléctrónicas pueden permitir el diseño de ventanas inteligentes que se oscurezcan según la intensidad de la luz, o pinturas que iluminen una habitación de manera uniforme, etc. Pero, como siempre que aparece algo nuevo, las aplicaciones que utilicen al máximo el potencial del material están aún por llegar. Entre las propuestas más llamativas en diverso grado de realización están:

a) Biomedicina: se ha comprobado que el óxido de grafeno rodea a las bacterias y perfora sus membranas impidiéndoles sobrevivir. Recubrir los instrumentos de quirófano con este óxido limitaría la necesidad de antibióticos en las operaciones. La unión de material biocompatible y muy buen conductor ha dado lugar a propuestas de electrodos basados en grafeno para implantar en el cerebro y ayudar al control motriz de personas con daños cerebrales.

b) Membrana: las membranas de óxido de grafeno bloquean de manera perfecta el paso de líquidos y gases. En Manchester se investiga en procesos de filtrado de aguas y desalinización de interés especial para el tercer mundo.

c) Sensores: Las propiedades del grafeno como membrana y su capacidad de deformarse ante el impacto de moléculas individuales permiten diseñar sensores de tamaño micrométrico capaces de detectar gases nocivos a nivel molecular.

69.4 ¿Algo ya en el mercado?

Quizás la aplicación más inmediata ya en curso está en su uso en pantallas táctiles flexibles. Las pantallas actuales están basadas en un cristal bañado en un óxido de indio-estaño, una sustancia transparente y conductora de tipo cerámico. El indio es un elemento escaso y muy caro, y los dispositivos se rompen con facilidad. El óxido de grafeno permite fabricar pantallas ligeras, elásticas e irrompibles a precios competitivos. El primer prototipo se realizó en septiembre de 2014 en un consorcio entre la Universidad de Cambridge y la empresa inglesa "Plastic Logic" y hay rumores de que una empresa china (Moxi) ha producido ya un teléfono enrollable como un reloj (ganando la carrera a Samsung). El anuncio es que van a poner a la venta en este año 2016 unos 100.000 ejemplares en el mercado chino a un precio aproximado de 700 euros. Los dispositivos pesan 200 g. Si no es ahora, será más tarde.

En el mercado hay también varios productos que integran el grafeno en su fabricación desde raquetas de tenis y esquíes (HEAD) ruedas de bicicletas de carreras (Vittoria) y de coches (Qingdao Sentury Tire), cascos para ciclistas (Catlike), etc. [1]

69.5 Unificación de varias ramas de la física.

A principios del siglo XX la física conoció dos grandes revoluciones: La relatividad general y la mecánica cuántica. En aquéllos años, los físicos eran físicos sin apellidos: Feynman, Landau, Wigner, Einstein, Dirac, trabajaron en problemas que hoy estarían clasificados en compartimentos estancos como teoría cuántica de campos, relatividad, física estadística, materia condensada. Debido a sus propiedades físicas, el grafeno constituye un ejemplo de unificación de ramas diversas de la física con el enriquecimiento que esto supone para el avance conceptual. Es una membrana (materiales blandos), rígida y elástica (teoría de elasticidad), impermeable (física-química), electrónica (ingeniería física) y relativista (teoría cuántica de campos). Además las corrugaciones que presentan las muestras se pueden modelizar con técnicas de teoría cuántica de campos en espacios curvos como las utilizadas en cosmología y "astropartículas".

69.6 Futuro

La síntesis del grafeno ha abierto la puerta a la de otros materiales bidimensionales. Aparte de la síntesis de los materiales de la misma columna que el carbono en la tabla periódica (fosforeno, siliceno, germaneno) se han sintetizado materiales compuestos bidimensionales con propiedades complementarias a las del grafeno. De particular interés son los llamados dicalcogenuros y el nitruro de boro. Apilando capas de distintos materiales se obtienen sólidos tridimensionales (llamados materiales de Van der Waals) funcionalizados para que tengan las propiedades deseadas. Inspirados en la física del grafeno, se han identificado recientemente materiales con propiedades electrónicas parecidas (sus portadores son también partículas sin masa) en tres dimensiones espaciales, los llamados fermiones de Weyl.[2]

Terminaremos con una advertencia y una recomendación: Saber más sobre el grafeno es fácil. Internet está plagado de información (sobre todo en inglés) de índole variada. Algunas páginas contienen exageraciones atractivas - y a veces incorrectas - sobre las futuras aplicaciones. El grupo de Manchester mantiene una página de divulgación actualizada y fiable donde acudir en caso de duda. [3]

Notas:
[1] (Esta información está sacada de la página http://www.graphene-info.com/graphene-products).
[2] (Un artículo de divulgación excelente sobre estos materiales aparecerá en breve (2016/2017) en la Revista de la Real Sociedad de Física).
[3] (Grupo de Manchester http://www.graphene.manchester.ac.uk).

María A. H. Vozmediano
Doctora en Física
Investigador Científico en el Instituto de Ciencia de Materiales de Madrid

Capítulo 70
¿Qué es el efecto Coriolis?
(Por Julio Gutiérrez Muñoz)

En el capítulo 36, epígrafe 36.2, analizamos el principio de relatividad, el cual establece que el estado de movimiento de un objeto no es absoluto, sino que depende del estado de movimiento del observador, es decir, la velocidad de un objeto es relativa al Sistema de Referencia (SR); si alguien camina por un vagón de un tren en marcha, la velocidad medida por otro pasajero no es la misma que mediría un observador situado en reposo respecto de la vía. En el escenario de la Relatividad Especial, las aceleraciones y, en consecuencia, las fuerzas tampoco son invariantes, es decir, dependen también del SR. Sin embargo, en un escenario de velocidades pequeñas frente a la velocidad de la luz, los cambios en las fuerzas y aceleraciones, al cambiar de SR, siempre que manejemos sistemas de referencia inerciales (abreviados por SRI), es decir, no acelerados, son despreciables y se puede utilizar la Relatividad de Galileo. En caso de utilizar sistemas no inerciales (abreviados SRNI), el panorama se complica enormemente. El movimiento de un objeto a velocidad constante en módulo y dirección en un SRI puede verse animado de fuertes aceleraciones en el seno de un SRNI, lo cual es contradictorio con los valores nulos que mostrarían los acelerómetros solidarios con el objeto.

Antes de continuar, recomendamos la lectura del epígrafe 37.2, a fin de recordar la naturaleza de las fuerzas ficticias o de inercia.

70.1 Aceleración de Coriolis

Un SRNI típico de los libros de texto consiste en una plataforma circular horizontal −a fin de no hacer intervenir la fuerza de la gravedad− que gira a velocidad angular constante, ω, y sobre la que pueden moverse los objetos sin rozamiento alguno. Las imágenes que siguen son suficientemente ilustrativas y apenas necesitan ser comentadas. La figura 1 representa, desde el punto de vista de un observador no inercial solidario con la plataforma giratoria y situado en el borde la misma, la trayectoria que sigue un objeto lanzado por él con velocidad inicial v_0 hacia el centro de la plataforma. El objeto, en vez de seguir la línea del diámetro, como era de esperar, traza una curva cerrada que lo lleva fuera del sistema. ¿Qué ha pasado? Al ser lanzado desde el interior del SRNI, está animado inicialmente de la velocidad de giro correspondiente al punto de partida, la cual, en el borde de la plataforma, tiene el valor:

$$v_C = \omega R \qquad (70\text{-}1)$$

siendo ω y R la velocidad angular y el radio de la plataforma, respectivamente. Esa velocidad va cambiando de dirección mientras el objeto está en la mano del lanzador

pero, en el momento del lanzamiento, v_C se mantendrá con su dirección de salida invariable, tangente al borde de la plataforma en ese instante. Es decir, vista desde un SRI, en el que el centro de la plataforma permanece en reposo, la velocidad del objeto tiene dos componentes constantes, v_0 en la dirección del diámetro de la plataforma y v_C en la dirección perpendicular. Para un observador inercial, como muestra la figura 2, la trayectoria es una línea recta, inclinada respecto la dirección inicial del diámetro de la plataforma, recorrida a velocidad constante, de valor:

$$v_T = \sqrt{v_0^2 + v_C^2} \qquad (70\text{-}2)$$

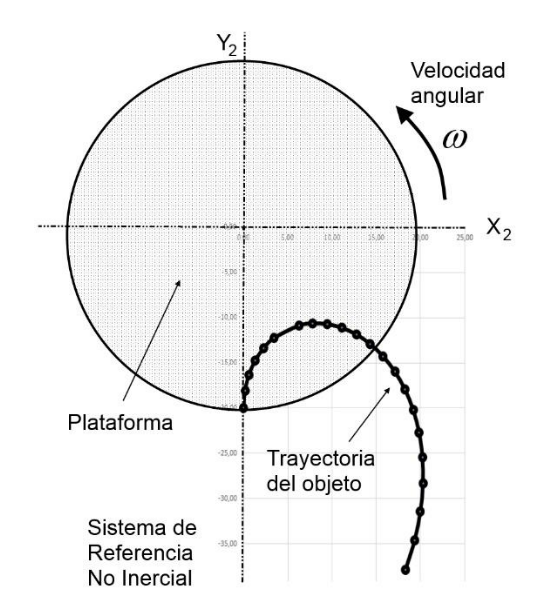


Figura 1: *Trayectoria de un objeto que parte del borde de la plataforma, animado con velocidad inicial de dos componentes: la de dicho borde y la del lanzamiento hacia el centro de la plataforma; visto por un observador no inercial situado en el centro de la plataforma del sistema de referencia en rotación.*

Figura 2: *El mismo objeto de la figura 1 visto desde un sistema de referencia inercial, en el que el centro de la plataforma está en reposo.*

Como se puede apreciar, en ambas figuras las distancia del centro de la plataforma a los puntos que indican la posición del objeto, en cada instante, son idénticas.

Esa aceleración aparente en el SRNI debe ser introducida como resultado de una fuerza ficticia denominada de Coriolis; la curvatura de la trayectoria recibe entonces el apelativo de **efecto Coriolis**. Un cálculo que, aunque no muy complicado, se sale del esquema de este libro, nos dice que, en cada instante, el valor vectorial de la aceleración ficticia de inercia, para un objeto animado de velocidad $\boldsymbol{v_T}$, en un SRNI, ligado a dicho sistema, que gira a velocidad angular constante, $\boldsymbol{\omega}$, tiene dos componentes y viene dada por:

$$\boldsymbol{a_{inercia}} = \boldsymbol{\omega} \text{ x } (\boldsymbol{\omega} \text{ x } \boldsymbol{r}) - 2\, \boldsymbol{\omega} \text{ x } \boldsymbol{v_T} \qquad (70\text{-}3)$$

siendo $\boldsymbol{r}$ la distancia vectorial desde el centro de la plataforma a la posición del objeto. En un caso general más complejo, como la del SRNI que supone la Tierra, $\boldsymbol{r}$ será la distancia vectorial desde el eje terrestre. Por convenio, a $\boldsymbol{\omega}$ se le asigna carácter vectorial en la dirección del eje de giro y sentido dado por el avance de un hipotético tornillo, cuya cabeza gira como lo hace la plataforma. El primer término de la ecuación anterior es la llamada **aceleración ficticia centrífuga**; el segundo término es la **aceleración de Coriolis**:

$$\boldsymbol{a}_{Coriolis} = -\,2\,\boldsymbol{\omega} \times \boldsymbol{v}_T \qquad (70\text{-}4)$$

Estas operaciones entre dos vectores, denotadas por el símbolo x, son denominadas **producto vectorial**, el cual indica que el vector resultante es perpendicular al plano formado por los dos vectores que se multiplican. Así, la aceleración de Coriolis, de módulo 2 $\omega v_T \operatorname{sen}\alpha$, siendo α el ángulo que forman los vectores $\boldsymbol{\omega}$ y $\boldsymbol{v}_T$, es, en todo instante, perpendicular tanto al eje de giro como a la velocidad del objeto, consecuentemente, esta última cambia de dirección en el SRNI pero nunca podrá cambiar en módulo. De esta forma, si el objeto tiene velocidad paralela al eje de giro, no presenta el efecto Coriolis, que es máximo cuando la velocidad es perpendicular el eje de giro. En otras palabras, la expresión anterior puede enunciarse diciendo que solo la componente de la velocidad del objeto perpendicular al eje de giro del SRNI da lugar al efecto Coriolis.

En el caso que nos ocupa, en el cual la plataforma permanece horizontal, $\boldsymbol{v}_T$ y $\boldsymbol{\omega}$ son perpendiculares entre sí durante todo el movimiento y el módulo será 2 ωv_T.

La aceleración centrífuga es diferente, solo viene dada por la posición del objeto en el SRNI respecto del eje de giro; no interviene su velocidad, solo la de rotación del SRNI. Como producto vectorial en cadena de tres vectores, perpendiculares entre sí, para casos de ω constante, es un vector en la dirección de $\boldsymbol{r}$ y módulo $\omega^2 r$. No vamos a hablar de su sentido, pues puede estar dirigida hacia el eje de giro o en sentido contrario. Seguramente el lector se habrá sorprendido con esta afirmación. El análisis de la naturaleza de esta fuerza ficticia, aunque parezca mentira, es muy complicado, tanto, que el 99 % de los textos que hablan de ella hacen un estudio erróneo; el resto la mencionan de pasada. Solo algunos distinguen entre la aceleración centrífuga ficticia y la real, debida a las ligaduras de los objetos con el SRNI, pero sin dar con la verdadera solución. En honor a la verdad, debemos decir que no hemos encontrado, hasta hoy, un solo texto que dé una solución correcta; incluso Wikipedia, como otros muchos, confunde la fuerza centrífuga real con la Tercera ley de Newton, aplicada a sistemas con fuerza centrípeta, por tener sentido contrario a esta última, lo cual es completamente absurdo. Efectivamente, si fuera una fuerza debida a la Ley de Acción y Reacción, debería estar aplicada **por** el objeto que sufre la fuerza centrípeta **sobre** el sistema no inercial, mientras que, en la realidad, es el propio objeto el que está sometido a las dos.

El hecho de que la aceleración de Coriolis no sea propia del movimiento inherente de un SRNI respecto de un SRI, sino de la cinemática del objeto, induce a muchos autores a denominarla "***aceleración complementaria de Coriolis***".

A modo de resumen, podemos decir que el efecto Coriolis no es sino la manifestación de la conservación del momento lineal –cantidad de movimiento–, así

como de su resultante giratoria –el momento angular– vista desde sistemas de referencia no inerciales animados de velocidad de rotación. El efecto Coriolis se hace notar en los lanzamientos de proyectiles, los movimientos sobre la superficie terrestre y, sobre todo, en los movimientos de las masas de aire de la atmósfera, las cuales, en vez de seguir la dirección perpendicular a las isobaras, como predice la Física de Fluidos, terminan por realizar movimientos giratorios similares a los de la Figura 1. Analicemos algunos casos particulares debidos a la rotación terrestre.

70.2 Efecto Coriolis en el SRNI de la Tierra.

Existen dos casos muy característicos de la aceleración de Coriolis.

Caso primero: Un objeto que se deja caer, sin velocidad inicial, desde una torre a cierta altura, *h*, sobre la superficie de la Tierra, está animado de una velocidad inicial de giro hacia el Este mayor que la velocidad de giro de la base de la torre. Efectivamente, la velocidad de giro de la Tierra, ***w***, es de 2π radianes en 24 horas, sea cual sea la altura o la latitud del lugar, por consiguiente, la velocidad lineal de cualquier objeto al ser abandonado a una altura *h* es de:

$$v_{0h} = \omega\,(R+h)\,\cos\theta \qquad (70\text{-}5)$$

siendo *R* el radio de la Tierra en el lugar del experimento, *h* la altura desde la base de la torre y θ el ángulo que determina la latitud. Sin embargo, la base tiene una velocidad de giro de:

$$v_0 = \omega R \cos\theta \qquad (70\text{-}6)$$

Estas expresiones nos dicen que el efecto es máximo en el ecuador ($\theta = 0º$) y nulo en los polos ($\theta = 90º$). Si despreciamos el rozamiento con el aire, mientras el objeto cae, este tiene una velocidad hacia el Este algo mayor que la base de la torre, por lo cual impactará en el suelo alejado de la misma ¡hacia el Este! Ahora bien, no podemos determinar la desviación calculando cuánto se ha desplazado el objeto respecto de la base de la torre durante el tiempo de caída utilizando la velocidad diferencial inicial entre la cima y la base de la torre, vista desde un SRI, pues, a medida que el objeto cae, la dirección de v_{0h}, que permanece invariable en módulo, dirección y sentido¸ y la velocidad de la base de la torre dejan de ser paralelas. El cálculo es algo complicado y se sale del esquema de este estudio.

No obstante, sí podemos hacer el cálculo de forma mucho más simple en el SRNI Tierra –el desvío hacia el Este debe ser igual en ambos sistemas de referencia–. Según la expresión (70-4), el valor de la aceleración de Coriolis es variable pues, en su caída, el objeto va adquiriendo velocidad con componente perpendicular al eje de giro:

$$a_{Coriolis} = 2t\,\omega\cos\theta \qquad (70\text{-}7)$$

Expresión que, sucesivamente integrada y teniendo en cuenta el tiempo de caída obtenido de la altura recorrida en caída libre sin velocidad inicial,

$$h = \frac{1}{2}\, g\, t^2 \qquad (70\text{-}8)$$

nos proporciona:

$$v_{hacia\ el\, este} = \omega\, g\, t^2 \cos\theta \qquad (70\text{-}9)$$

$$s_{hacia\ el\, este} = \frac{1}{3}\omega\, g\, t^3 \cos\theta = \frac{2}{3}\,\omega\, h \cos\theta \sqrt{\frac{2h}{g}} \qquad (70\text{-}10)$$

Para una torre de 100 metros, a la latitud de Madrid, 40° 24′ 59″ N, la desviación hacia el Este sería de 1,66 cm.

Caso segundo: Sea un móvil que parte de cierto punto y viaja sobre la superficie de la Tierra, cambiando de latitud. Si viaja desde el Ecuador hacia el Norte o hacia el Sur, aumentando su latitud relativa, la velocidad lineal de partida es mayor, hacia el Este, que la necesaria para mantener el giro terrestre, por consiguiente, el objeto tenderá a desviarse hacia el Este. Si por el contrario viaja desde el Norte o el Sur hacia el Ecuador, disminuyendo su latitud relativa, el resultado será una desviación hacia el Oeste. En estos casos, la componente de la velocidad perpendicular al eje de rotación terrestre es pequeña y viene dada por:

$$v_{\perp} = v_{objeto}\ \mathrm{sen}\theta \qquad (70\text{-}11)$$

Por ello, en este supuesto, la aceleración de Coriolis es nula en el Ecuador –la velocidad del objeto es paralela a $\boldsymbol{\omega}$–, aumenta con la latitud y es máxima en los polos; a medida que nos acercamos a los polos, el cambio de velocidad lineal de rotación es mayor para la misma variación del ángulo de latitud.

70.3 Efecto Eötvös.

Existen otras componentes significativas verticales del efecto Coriolis. Los objetos que viajan a cierta velocidad hacia el Este sobre un paralelo terrestre tendrán aceleración ficticia dirigida hacia alturas mayores (parecerán más ligeros), mientras que los que lo hagan hacia el Oeste tendrán aceleración hacia abajo (parecerán más pesados).

Examinemos el efecto para un objeto que se mueve a lo largo del Ecuador a nivel del mar con velocidad v_T = 100 km/h. La ecuación (70-3) nos dice que en el Sistema de Referencia Tierra, existen dos aceleraciones ficticias, ambas en la dirección del radio terrestre, R_T:

1) la correspondiente al primer miembro del lado derecho, la aceleración centrífuga, con sentido contrario a la aceleración de la gravedad y de valor

$$a_{centrífuga} = \omega^2 R_T = 0{,}034 \text{ m/s}^2 \qquad (70\text{-}12)$$

2) la correspondiente al segundo miembro del lado derecho, la aceleración de Coriolis, de valor absoluto

$$a_{Coriolis} = 2\,\omega\,v_T \;=\; 0{,}004 \text{ m/s}^2 \qquad (70\text{-}13)$$

con sentido contrario a la gravedad cuando el objeto se mueve hacia el Este y del mismo sentido cuando lo hace hacia el Oeste. Es decir, el objeto pesa menos en el primer caso y un poco más en el segundo. Sin embargo, la aceleración centrífuga siempre ayuda a los objetos a pesar algo menos que lo predicho por la Ley de Newton de Gravitación, que daría un valor para el Ecuador:

$$a_{gravedad} = \frac{\mathrm{G\,M}_T}{R_T^2} = 9{,}82 \text{ m/s}^2 \qquad (70\text{-}14)$$

En una latitud superior como la de Madrid, 40º 24’ 59”, hemos de tener en cuenta el efecto de achatamiento terrestre, que aumenta la atracción gravitatoria, y la altitud que la disminuye, mientras que la latitud afecta tanto a las aceleraciones centrífuga y de Coriolis con un término multiplicativo en $\cos^2\theta$. Los valores correspondientes son:

$$a_{gravedad} = 9{,}82 \text{ m/s}^2 \;\;;\;\; a_{centrífuga} = 0{,}014 \text{ m/s}^2;\;\; a_{Coriolis} = 0{,}002 \text{ m/s}^2$$

Estos fenómenos se conocen como efecto Eötvös. Las componentes verticales de la aceleración de Coriolis son máximas en el Ecuador y van disminuyendo a medida que nos acercamos a los polos, al igual que las contribuciones de la aceleración centrífuga.

Julio Gutiérrez Muñoz
Doctor en Física
Catedrático de Universidad de Física Atómica, Molecular y Nuclear, jubilado.

Capítulo 71
Arquímedes y la flotabilidad de los cuerpos sumergidos
(Por Mario Sánchez Sanz)

El "eureka" de Arquímedes es, con permiso de la manzana de Newton, el falso mito asociado a la ciencia más reconocible por todos nosotros. La imagen del joven matemático corriendo desnudo por la ciudad griega de Siracusa, con la falsa corona de oro del rey Hierón en la mano, es suficientemente sugerente como para dejar que la realidad estropee una buena historia. Dejando de lado la anécdota, la contribución de Arquímedes fue de tal importancia que cualquier estudiante de secundaria es capaz de recitar de memoria el principio de Arquímedes:

"la fuerza de flotabilidad que aparece en un cuerpo sumergido es igual al peso del volumen de fluido que el cuerpo desaloja".

Enunciado de esta manera es probable que nos preguntemos a nosotros mismos sí realmente entendemos su significado. Afortunadamente, todos hemos experimentado esa fuerza de flotabilidad de la que habla Arquímedes cuando nos bañamos en una piscina o en el mar. Si recordamos el esfuerzo que nos requiere levantar un objeto pesado en el agua e intentamos repetir esa acción fuera de ella, nos daremos cuenta rápidamente que la sensación de ligereza del primer ejercicio no se debía a nuestras duras sesiones de gimnasio, sino a la ayuda que el fluido nos estaba prestando. Barcos, submarinos, boyas y un largo etcétera explican su funcionamiento a través del principio de Arquímedes.

A pesar de no ser mencionado, el concepto fundamental que subyace implícitamente en el enunciado del principio de Arquímedes es el de presión hidrostática. La presión se define como una fuerza por unidad de superficie y tiene unidades de Pascales o N/m^2. Bajo un campo de presión constante, es posible calcular la fuerza que esa presión ejerce sobre una superficie de área S multiplicando el valor de la presión por S. Para campos de presiones variables, el cálculo de la fuerza implica, generalmente, el uso del cálculo integral y puede llegar a ser complicado en cuerpos con superficies irregulares.

Aunque nos hayamos acostumbrado a ella y no la notemos, la presión forma parte de nuestra vida diaria. Cuando caminamos por la calle y nos dirigimos a nuestros quehaceres diarios, nos olvidamos que nos movemos en el interior de un fluido, el aire. Al nivel del mar, cualquiera de nosotros siente una presión constante, que conocemos como presión atmosférica, generada por el peso de la ancha capa de aire que nos separa del espacio exterior. Si asumimos que el aire está quieto, es posible calcular la magnitud de esa presión multiplicando la densidad del aire (1.1 kg/m^3 a 20 ºC) por la aceleración de la gravedad (9.81 m/s^2) y por la anchura de la atmósfera (unos 9.5 km, aproximadamente) para obtener un número que se acerca a 101300 N/m^2 o 1 atmósfera

de presión. Cuanto mayor sea el espesor de la capa de aire o mayor sea la densidad del aire, mayor será la presión que tenemos que soportar. Lo contrario ocurre cuando disminuimos la densidad del aire o disminuye el espesor de la capa de aire que tenemos sobre nosotros, como bien saben los montañeros que sufren de “mal de altura” debido a la disminución de la presión atmosférica al ascender por encima de los 4000 o 5000 metros. La densidad del fluido es fundamental a la hora de determinar como cambia la presión. Como acabamos de describir, para encontrar variaciones notables en la presión atmosférica necesitamos variar la anchura de la capa de aire sobre nosotros en cantidades cercanas al kilómetro. Eso no ocurre si estudiamos como cambia la presión en el agua a medida que nos sumergimos en ella. Aunque los principios que explican los cambios de presión son los mismos, el incremento de presión con la profundidad es mucho más rápido en el agua debido a que su densidad ρ=1000 kg/m^3 es casi 1000 veces superior a la densidad del aire ρ=1.1 kg/m^3. Así, cada vez que nos sumergimos 10 metros en el mar, la presión aumenta en aproximadamente una atmósfera [1].

Para entender el principio de Arquímedes vamos a realizar un sencillo experimento mental. Para ello vamos a sumergir, en agua, un cubo de Rubik. Este juguete, en el que muchos hemos invertido muchas más horas de las que deberíamos en épocas de exámenes, no es más que un prisma cuadrado con una longitud de lado “L” conocida. Para estudiar en que consiste la fuerza de la flotabilidad, sumergimos el cubo en un fluido de tal forma que dos de sus caras permanezcan paralelas a la superficie. Para medir la profundidad a la que colocamos el cubo, situamos nuestro sistema de referencia en la superficie de agua de forma que la profundidad la definimos como la distancia vertical medida desde la superficie del agua hasta la cara superior del cubo. Como ya hemos visto anteriormente, la presión aumenta a medida que sumergimos el cubo a una mayor profundidad, de forma que si la distancia entre la cara superior del cubo y la superficie del agua es “h”, la presión a la que se ve sometida esa cara del cubo será $p = \rho \cdot g \cdot h$. Esa presión induce una fuerza vertical $F_s = -\rho g h L^2$, donde el signo negativo indica que la fuerza tiende a desplazar el cubo hacia aguas más profundas. Si repetimos el ejercicio sobre la cara inferior, la presión en esa cara será $p = \rho \cdot g \cdot (h + L)$y la fuerza asociada $F_i = \rho g(h + L)L^2$, en este caso con signo positivo porque esa fuerza trata de desplazar al cubo hacia la superficie del fluido. Sumando ambas fuerzas, obtenemos la fuerza vertical neta inducida por el campo de presiones sobre nuestro cubo de Rubik para dar $F = F_s + F_i = \rho g L^3$, magnitud positiva que representa el peso del volumen de agua desalojado por el cubo, cuyo volumen es L^3.

La fuerza neta F que acabamos de calcular, conocida también como la fuerza de flotabilidad, es independiente de la profundidad a la que coloquemos nuestro cubo y será siempre la misma si el cubo está completamente sumergido. Obviamente, para saber si eso va a ser así, tenemos que tener en cuenta que el cubo tiene un peso W. Si consideramos que el material del que está hecho el cubo tiene una densidad ρ_c, su peso lo podemos expresar como $W = -\rho_c g L^3$donde el signo negativo indica que la esa fuerza intenta desplazar al cubo hacia profundidades mayores. Teniendo en cuenta que sobre el cubo únicamente actúan estás dos fuerzas, la segunda ley de Newton nos permite determinar la dirección hacia la cual se desplazará nuestro prisma en función del signo del resultado de la suma de nuestras dos fuerzas $F + W = (\rho - \rho_c) g L^3$:

1) Si la suma de las dos fuerzas resulta ser idénticamente nula, nuestro cubo se encuentra en equilibrio y se quedará, hasta el final de los tiempos, en la misma posición

en la que lo hemos colocado inicialmente. Esto ocurre únicamente si las densidades del fluido y del cuerpo que hemos sumergido son idénticas $\rho = \rho_c$.

2) Si la suma es positiva, el resultado indicaría que la fuerza de flotabilidad F es mayor que el peso y el cubo se desplazaría hacia la superficie del agua, donde quedaría parcialmente sumergido hasta que la fuerza de flotabilidad igualara al peso. Este caso representa lo que ocurre con los barcos y lo analizaremos con más detalle un poco más abajo.

3) Si la suma es negativa, la fuerza de flotabilidad no es suficiente para compensar al peso del cubo y éste se hundiría irremediablemente hasta llegar al fondo del mar.

El caso número 2, en el que el cubo queda parcialmente sumergido en la superficie del agua, podemos utilizarlo para entender por qué flotan los barcos. Claro, hasta el momento ya sabemos que cuando el cubo se encuentra completamente sumergido, la fuerza de flotabilidad F se calcula tal y como hemos indicado más arriba. Si la densidad del material del que está hecho el cubo es más pequeña que la densidad del fluido en el que el cubo está inmerso $\rho > \rho_c$, el cubo asciende hasta alcanzar la superficie donde permanece parcialmente sumergido. Para obtener la proporción del cubo que queda bajo el agua solo tenemos que igualar las dos fuerzas y obtener el volumen del cubo que debería estar sumergido para que la fuerza de flotabilidad igual al peso, que resulta ser $\frac{\rho_c}{\rho} L^3 < L^3$. Cuanto mayor sea la densidad del fluido respecto a la densidad del cubo, menor será el volumen del cubo que queda sumergido.

El resultado que hemos obtenido anteriormente se puede generalizar para cuerpos con geometrías más complejas que la del cubo de Rubik que hemos usado nosotros. Es posible demostrar, con ayuda del cálculo integral y del teorema de Gauss, que la fuerza de flotabilidad se puede expresar, de forma general, como $F = \rho g V$, donde V representan el volumen del cuerpo que está sumergido en el fluido que consideremos. La construcción de barcos, desde la antigüedad, se basa en esta idea. El diseño trata de buscar minimizar el peso del barco, usando para ello materiales lo más ligeros posibles, maximizando la flotabilidad con un diseño apropiado del casco del barco. Obviamente, no es lo mismo el diseño de un barco que se mueve a baja velocidad que uno pensado para desplazamientos rápidos, donde el casco se eleva para disminuir la superficie de contacto con el agua reduciendo, de esta forma, la resistencia que el líquido presenta al desplazamiento del barco. En barcos lentos, como pueden ser los barcos de transporte de mercancías, el buque navegando desplaza prácticamente el mismo volumen de agua sumergida que en parado.

Una de las aplicaciones más espectaculares del principio de Arquímedes es, sin duda, el submarino. Concebido por Isaac Peral en 1884, el primer submarino construido fue botado en 1888 consiguiendo, tras unos pocos ensayos, simular el ataque a un navío durante la noche volviendo a puerto sin ser detectado. Unos años más tarde, su diseño fue mejorado por John Phillip Holland sustituyendo las baterías que Peral instaló como sistema de propulsión por un motor de combustión interna.

El principio de funcionamiento del submarino es conceptualmente sencillo. Está compuesto por un doble casco que forma una cámara que puede llenarse de aire o agua en función de si se quiere que el submarino descienda o ascienda. Cuando el submarino

se encuentra en flotación, se introduce agua dentro de esa cámara aumentando, así, su peso. Puesto que la flotabilidad depende del volumen del casco interior, a medida que se llena de agua la cámara conseguimos un desplazamiento controlado hacia aguas más profundas. Por el contrario, para ascender, se vacía de agua la cámara formada por los dos cascos llenándola de aire comprimido. Con esta sencilla maniobra, se consigue disminuir el peso y aumentar la flotabilidad del submarino. A medida que el submarino desciende a mayores profundidades, la presión que debe soportar el casco es mayor. En el diseño de submarinos se define la máxima profundidad de operación y la profundidad de colapso. La primera marca la máxima profundidad a la que el submarino podría sumergirse para operación normal mientras la segunda es la profundidad a la que el casco del submarino fallaría por colapso estructural. Los submarinos atómicos modernos de la clase Seawolf son capaces de alcanzar profundidades de operación de 490 metros [2] (47.4 atmósferas de presión) y profundidades de colapso de 730 metros (70 atmósferas de presión).

En el año 2012, el director de cine James Cameron descendió, con la ayuda de un submarino especialmente diseñado para soportar altas presiones, hasta los 11000 metros de profundidad. El batiscafo que le protegía soportó cerca de 1064 atmósferas de presión.

De forma rutinaria, ciertos mamíferos marinos alcanzan 1500 metros de profundidad en busca de alimento. Como hemos aprendido anteriormente, el balance entre la flotabilidad y el peso determina la profundidad a la que un cuerpo sumergido encuentra el equilibrio. Por una razón u otra, peces y mamíferos marinos se ven obligado a modificar la profundidad a la que nadan, acción que, en principio, implicaría un esfuerzo muscular y un consumo de energía. Para entender este hecho, recurro una vez más al ejemplo de la piscina para que el lector experimente en su propio cuerpo el esfuerzo que es necesario realizar para mantener la profundidad de inmersión. Claro, si no podemos cambiar nuestro peso y no podemos modificar nuestra flotabilidad, la única manera de permanece estacionarios a una distancia de la superficie es nadando, es decir, realizando una fuerza que compense el exceso de peso, que nos llevaría al fondo, o de flotabilidad, que nos llevaría a la superficie.

Los peces y mamíferos marinos han desarrollado un órgano extraordinario que permite modificar su densidad e igualarla a la del agua que les rodea para evitar malgastar energía en mantener su profundidad. La vejiga natatoria es un órgano de flotación con paredes flexibles, que se sitúa en la columna vertebral y que permite modificar la densidad del pez introduciendo aire en ella cada vez que acceden a la superficie del agua. Para modificar su densidad, expulsan poco a poco el aire de la vejiga hasta llegar a la profundidad deseada. De esa manera, solo es necesario hacer un esfuerzo muscular para regresar a la superficie del agua, donde llenaría de nuevo su vejiga para volver a sumergirse.

Notas:

[1] Claro, como la densidad del agua es ρ=1000 kg/m3, para que la presión aumente en una atmósfera, el espesor de la columna de agua sobre nosotros debe ser $z = 101300/(1000 \cdot 9.8) = 10.33m$

[2] Federation of American Scientists (8 December 1998). "Run Silent, Run Deep". *Military Analysis Network*. Retrieved 10 May 2010.

Mario Sánchez Sanz
Doctor en Ingeniería Matemática
Profesor Titular, Universidad Carlos III de Madrid

Capítulo 72
¿Un futuro para la ciencia? Una visión desde la química.
(Por Bernardo Herradón)

Desde que nos levantamos hasta que nos acostamos, estamos rodeados de ciencia. Como se indicará más adelante, los beneficios que obtenemos de la ciencia (y sus aplicaciones, la tecnología) superan con mucho a sus posibles riesgos o perjuicios. También hay que tener en cuenta que la mayoría de los aspectos negativos relacionados con la ciencia y la tecnología se deben a su mal uso.

En este capítulo nos aventuramos a hacer pronósticos sobre el futuro de la ciencia, especialmente de la química, basándonos en su evolución reciente, sin olvidar alguna llamada de atención sobre aspectos que pueden hacer peligrar su progreso.

72.1 ¿Habrá ciencia en el futuro?

Esta parece una pregunta de fácil respuesta, ¿cómo no va a existir la ciencia?, ¡con la inmensa cantidad de comodidades y avances que nos proporciona! Sin embargo, hay signos preocupantes que no hacen ser muy optimista sobre el futuro ni tan contundente en esta respuesta. Estas señales son globales, pero se acentúan en países con poca tradición científica, como España. Algunos aspectos preocupantes son:

72.1.1 Baja percepción social de la ciencia.

Muchas personas no son capaces (o no quieren) de reconocer lo que la ciencia ha hecho por su calidad de vida; e, incluso se destacan aspectos negativos puntuales, tales como "la contaminación ambiental", "el miedo a los transgénicos", "la desconfianza sobre métodos para producir energía", etc.

72.1.2 Pobre cultura científica.

La sociedad actual aprecia ciertos valores; en los que, por desgracia, la cultura (sin calificativos) no ocupa un papel predominante. Algunos de los asuntos indicados en el párrafo anterior tienen que ver con la falta de cultura del receptor de la información.

72.1.3 El 'éxito' de las pseudociencias.

Como consecuencia de las situaciones comentadas en los dos párrafos anteriores, estamos invadidos por numerosas actividades que, sin base científica, nos "las venden" como científicas. Entre estas podemos citar la relacionadas con la medicina (con la homeopatía, los movimientos anti-vacuna como actividades más destacadas o la mal llamada 'medicina alternativa') [1], que supone un peligro para la salud individual y

colectiva; o falsas creencias (la astrología es un ejemplo significativo); o las religiones que se mezclan con la ciencia; entre otras.

72.1.4 El papel de los científicos.

Muchos científicos son responsables de la crisis por la que está pasando la ciencia. En este aspecto podemos destacar dos facetas: su resistencia a transmitir conocimiento (cultura científica) y la 'exageración' a la hora de destacar sus logros. En octubre de 2013, la revista The Economist publicó el extenso informe How Science goes Wrong sobre aspectos negativos para la ciencia, como la irreproducibilidad de resultados, la retirada de artículos por mostrar datos incorrectos, etc.

72.1.5 La política científica.

Por desgracia, en algunos países como el nuestro, la cultura, la educación y la ciencia nunca son prioridades políticas [2]. Comentar que debido a la gravísima crisis económica que estamos padeciendo, las inversiones en estas áreas han disminuido considerablemente. Para la ciencia esta situación es catastrófica, especialmente en dos aspectos. Por un lado, está produciendo que muchos proyectos de investigación se tienen que paralizar (así como el mantenimiento de grandes equipos) y será mucho más costoso y trabajoso si se tienen que retomar en el futuro. Por otro lado, la esperanza de los jóvenes de realizar una carrera científica se está viendo frustrada.

Por lo tanto, vista la situación de la ciencia actualmente, nos podemos preguntar si merece la pena hacer predicciones de futuro sobre la ciencia, su desarrollo y sus aplicaciones. Sin embargo, seremos optimistas y pensaremos que las deficiencias apuntadas en los párrafos anteriores son coyunturales y que habrá ciencia en el futuro. Antes de ir a las predicciones, conviene recordar los beneficios de la ciencia para la humanidad.

72.2 Lo que la ciencia nos proporciona.

Desde los albores de la humanidad con la generación y control del fuego, hemos visto progresos considerables que han desembocado en la situación actual con una civilización tecnológicamente muy desarrollada, en la que gracias a la ciencia tenemos:

72.2.1 Una vida más larga.

El aumento de la esperanza de vida al nacer se ha duplicado en poco más de 100 años, principalmente debido a los avances médicos basados en desarrollos científicos en diversas áreas.

72.2.2 La vida es más saludable.

Monitoriza nuestra salud. Proporciona medicinas que curan nuestras enfermedades, piezas de recambio para nuestro cuerpo, palia dolores y achaques.

72.2.3 Agua pura y potable.

Gracias a una combinación de procesos químicos y químico-físicos, disponemos de agua que podemos beber, usar para nuestra higiene o regar nuestras plantaciones.

72.2.4 Cuidado del ganado y animales de compañía.

El aumento en la esperanza de vida también se hace patente el resto de animales, en particular, el ganado y los animales de compañía.

72.2.5 Más y mejores alimentos.

Disponemos de campos más productivos y un ganado más cuidado. Una vez producido el alimento, lo podemos conservar más tiempo en mejores condiciones.

72.2.6 Nos proporciona energía.

Calor en invierno, frescor en verano, electricidad para la iluminación, nos permite circular en vehículos, etc.

72.2.7 Nuestra vida cotidiana es más cómoda.

Con el uso de electrodomésticos, la iluminación, el transporte, etc.

72.2.8 Objetos de nuestra vida cotidiana.

Hace que nuestras ropas y sus colores sean más resistentes y atractivos; mejora nuestro aspecto con perfumes, productos de higiene y de cosmética; contribuye en la limpieza del hogar y de nuestros utensilios; ayuda a mantener frescos nuestros alimentos; y prácticamente nos proporciona todos los artículos que usamos a diario.

72.2.9 Facilita el ocio.

Actividades como el deporte, la jardinería, la lectura, escuchar música, etc. le deben mucho a la ciencia.

72.2.10 Estar a la última en tecnología.

Disponer del ordenador más potente y ligero, el teléfono móvil más versátil y con una batería duradera y ligera, el sistema más moderno de iluminación, el medio de transporte adecuado, el material para batir marcas deportivos; y muchas aplicaciones más.

72.2.11 “Alimenta” el espíritu.

Este aspecto no hay que desdeñarlo. El ser humano es la única especie del planeta que puede disfrutar del placer de aprender, ahondar en el conocimiento. Esta es una faceta en la que la investigación básica juega un papel fundamental.

72.3 ¿Es la ciencia predecible? ¿Deben los científicos hacer predicciones?

La respuesta a la primera pregunta es SI y NO.

72.3.1 Ciencia predecible.

La ciencia predecible permite avanzar lento pero seguro; esto no es negativo, pues contribuye a satisfacer la mayoría de las necesidades del ser humano. La ciencia predecible es el resultado del diseño basado en el conocimiento previo. Por poner un ejemplo. Imaginemos que necesitamos una sustancia química para tratar una enfermedad y que este fármaco solo supone una ligera mejora respecto a lo anteriormente conocido (esta mejora puede ser debida a muchos factores: más actividad biológica, mayor facilidad de absorción, más fácil dosificación, más estabilidad *in vivo*, etc.) y que, con los conocimientos actuales de la química, podemos prepararla con 'relativa' facilidad. La ciencia lo hace y se consigue un ligero avance usando ciencia predecible.

72.3.2 Ciencia no predecible.

La ciencia no predecible es en la que se obtienen resultados inesperados. Una vez que se ha confirmado que los resultados son reproducibles, los científicos deben explicar estos resultados. Este tipo de resultados son los que dan origen a las revoluciones científicas. Como ejemplo nos puede servir todos los resultados sorprendentes que se obtuvieron a finales del siglo XIX y principios del siglo XX, como los descubrimientos de los rayos X y de la radiactividad, las pruebas de la existencia del electrón, el efecto fotoeléctrico, los experimentos de bombardeo de láminas metálica con partículas a, etc. Todos estos hallazgos requirieron el desarrollo de nuevas teorías para explicarlos, dando lugar al nacimiento de la mecánica cuántica.

72.3.3 ¿Deben los científicos predecir?

Por otro lado, ¿deben los científicos predecir? La respuesta es SI. Como Whitesides ha expuesto, hay varias razones [3]:

a) planificar nuestro trabajo
b) alimentar nuestra curiosidad
c) por razones filosóficas, pues la ciencia y la tecnología son importantes elementos culturales en nuestro tiempo y sociedad
d) por razones éticas: ¿qué investigación no deberíamos hacer?
e) los científicos y la ciencia se deben implicar en los cambios sociales
f) la sociedad espera que los científicos especulemos con el futuro.

72.4 El futuro: ciencia básica.

Aunque estamos jugando a adivino, los pronósticos se basan en algunos avances científicos recientes que deberán desarrollarse en las próximas décadas. Aunque en el título de este apartado se menciona la 'ciencia básica', estamos de acuerdo con la frase de Pasteur de que "*no existen áreas de la ciencia a las que podamos llamar básica y aplicada, sino que existe la ciencia y sus aplicaciones*".

Sin embargo, como se discutirá en el apartado siguiente, la ciencia debería avanzar para satisfacer las necesidades del ser humano; pero siempre recordando que, históricamente, la inmensa mayoría de los dispositivos que usamos actualmente han surgido de investigaciones básicas que no pretendían ninguna aplicación práctica.

Actualmente en todas las ciencias se vive una situación ‘excitante’, con avances espectaculares recientes que deberían ser confirmados y desarrollados en las próximas décadas. A continuación se apuntan algunas áreas de desarrollo futuro en las cinco ciencias básicas.

72.4.1 Física.

El principal objetivo de la física es, nada más y nada menos, que entender las leyes que rigen el universo y expresarlas en forma matemática. La confirmación experimental de la existencia del bosón de Higgs y los hallazgos que se están produciendo en el LHC y en otros aceleradores de partículas suponen un avance considerable en el modelo estándar que intenta explicar la interacción entre la materia y la energía desde el punto de vista de las interacciones fundamentales (electromagnetismo, nuclear fuerte y nuclear débil); pero aún queda mucho por descubrir. Algunos de los mayores avances en física estarán relacionados con la teoría cuántica de la gravitación (y la detección del gravitón, la partícula que media en esta interacción), la materia oscura, la energía oscura y el origen del universo. El reciente anuncio de la detección de ondas gravitacionales también es un progreso muy importante para entender el origen y destino del universo.

72.4.2 Biología.

El objetivo de la biología no es menos importante que el de la física: entender qué es la vida y su evolución. La biología moderna se sustenta en dos aproximaciones científicas: la evolutiva y la molecular, frecuentemente unidas. En los últimos años hemos presenciado avances espectaculares en la filogenética de numerosos seres vivos, especialmente del ser humano, basado en los progresos de la biología molecular. Este tipo de investigaciones, en la frontera entre la biología y la paleontología, se seguirá desarrollando en las próximas décadas. Recientemente se han publicado avances impresionantes en la denominada biología sintética. Dejando aparte denominaciones tan llamativas (y falsas) como “que los científicos juegan a ser Dios”, esta área científica puede ayudarnos a entender los aspectos básicos de lo que constituye la vida y su evolución.

72.4.3 Geología.

En geología se seguirá avanzando en la comprensión de las erupciones volcánicas y en los terremotos, que pueden servir para su predicción. La investigación de objetos extraterrestres supondrá un tema de interés para las próximas generaciones de geólogos.

72.4.4 Matemática.

La mayoría de los matemáticos seguirán investigando en temas que, aparentemente, no tienen aplicaciones prácticas; sin embargo, como se ha demostrado a

lo largo de la historia, muchos desarrollos científicos en matemáticas han tenido aplicaciones en áreas científicas y tecnológicas en áreas muy distantes [4].

72.4.5 Química.

Finalmente, la química debería seguir cumpliendo el papel que lleva realizando desde hace más de 200 años: proporcionar los materiales que usamos en nuestra vida cotidiana (la química es la ciencia de las cosas cotidianas), lo que se discutirá más adelante. Sin embargo, la química también tiene que desarrollarse en algunos aspectos que establezcan de manera inequívoca sus fundamentos, (lo que ya ocurre en física). Este objetivo se logrará a través de la matematización de la química, lo que servirá para: a) establecer sus bases teóricas; b) interpretar más fácilmente los resultados; y c) aumentar el poder de predicción. Los avances en matemáticas, física y ciencias de la computación serán de gran ayuda para alcanzar estos objetivos. El desarrollo en el área frontera entre la química y las matemáticas también tiene implicaciones en la filosofía de la ciencia, pues puede ayudar a profundizar en las relaciones entre estas dos ciencias y la física (con las dos opciones extremas: el reduccionismo y el autonomismo).

72.5 El futuro: las aplicaciones de la química.

¿Qué debe hacer la ciencia, en general, y la química, en particular, en el futuro? La respuesta es fácil. Como en los dos últimos siglos: la química debe satisfacer las necesidades de la sociedad.

Aunque no sabemos cuales serán estas necesidades, podemos anticipar que, tal como está la sociedad actualmente, las áreas (muy amplias) que tendremos que atender son:

72.5.1 Energía.

Actualmente nuestra sociedad es consumidora en exceso de energía. Esto es consecuencia de nuestro progreso. La energía que consumimos procede principalmente de los combustibles fósiles (petróleo, gas natural y carbón). Los inconvenientes son graves: recursos limitados, no renovables, y contaminantes. Además, desde el punto de vista químico, quemar derivados del petróleo supone que estamos gastando miles de compuestos químicos que suponen las materias primas con las que fabricamos bienes de consumo, principalmente los plásticos y polímeros con infinidad de aplicaciones. El futuro de la energía depende del uso de fuentes renovables de energía (por ejemplo, la solar), que convertiremos en electricidad. La química está desarrollando procesos y materiales con lo que se puede aprovechar más eficientemente la energía solar y almacenar de manera adecuada energía eléctrica (pilas, baterías, supercondensadores, células de combustible, etc.).

72.5.2 Medio ambiente.

El deterioro medioambiental está estrechamente relacionado con nuestro consumo excesivo de energía. Si conseguimos los objetivos indicados en el apartado anterior, también contribuiremos a resolver el problema medioambiental. Es indudable que parte de la culpa de la alta contaminación ambiental se debe al uso excesivo y no adecuado de sustancias químicas. La química contribuirá implantando procesos

industriales que sean medioambientalmente más adecuado (la conocida como química verde y química sostenible), sustancias químicas menos perjudiciales e investigando métodos para eliminar contaminantes ambientales, como por ejemplo procesos que sean capaces de eliminar el dióxido de carbono o los óxidos de nitrógeno de la atmósfera.

72.5.3 Salud.

La química seguirá proporcionando compuestos biológicamente activos que se podrán usar como fármacos. También se desarrollarán biomateriales que servirán para reparar o reemplazar partes de nuestro cuerpo, un área de intenso desarrollo actualmente. Con sustancias químicas se estudiarán procesos biológicos con implicaciones en el desarrollo de enfermedades (la química como una herramienta en biomedicina).

72.5.4 Alimentos y agua.

La química seguirá contribuyendo a que los campos sean más productivos, mejoraremos la calidad de los alimentos, haciéndolos más seguros. Un aspecto importante es que tenemos que conseguir métodos de purificación de agua que sean transportables a cualquier lugar del planeta.

72.5.5 Materiales tecnológicos.

El futuro dependerá de tener instrumentos eficientes, pequeños y atractivos que sean útiles en nuestro trabajo, ocio y vida cotidiana. Dentro de estos desarrollos tenderemos a la miniaturización (la nanotecnología es el futuro y la química tiene mucho que aportar) en dispositivos electrónicos, mejores equipamientos deportivos y muchas más aplicaciones.

72.5.6 Aspectos sociales.

Además la ciencia del siglo XXI tiene que cumplir una función social, contribuyendo a las acciones de igualdad (de género y entre países). Concretamente, la química debe contribuir a la mejora de las condiciones de vida (salud, higiene, alimentación, agua) de los países en vías de desarrollo.

Para lograr todos estos objetivos, miles de químicos de todo el mundo están investigando activamente para obtener y estudiar sustancias químicas (materiales y fármacos) con lo que abordar los retos indicados en los párrafos anteriores. Algunas áreas de desarrollo actualmente son:

-Máquinas moleculares
-Cristales líquidos más resistentes y energéticamente más adecuados
-Materiales con óptica no lineal
-Electrónica molecular
-Interruptores moleculares (en electrónica o en computación)
-Ordenadores moleculares
-Materiales quimioluminiscentes (conversión de energía química en luz)
-Diodos emisores de luz
-Antenas de luz (conversión de energía lumínica en química, centros fotosintéticos artificiales)

-Materiales nanoestructurados
-Equipos pequeños para monitorización (ambiental, salud, etc.)
-Almacenamiento y transporte de hidrógeno
-Almacenamiento y transporte de electricidad
-Producción eficiente de energía (con el menor impacto medioambiental)
-Convertidores de energía (química/lumínica/eléctrica).
-Nanocápsulas para transporte de fármacos
-Biomateriales: inertes biológicamente o que se integren en el tejido (funcionales)
-Purificación y potabilización de agua
-Fármacos para tratar enfermedades neurodegenerativas
-Compuestos para estudiar procesos biológicos con aplicaciones en biomedicina (interacción de proteínas, ácidos nucleicos, rutas de señalización, inmunología, etc.)
-Fármacos para tratar enfermedades raras
-Fármacos para tratar enfermedades de países en vías de desarrollo

Y todo esto, de manera eficaz; lo que, en química, significa: con el mínimo impacto ambiental, con el mayor aprovechamiento energético, y lo más barato posible.

72.6 Ciencia y futuro.

En este breve capítulo nos hemos atrevido a pronosticar algunos aspectos del futuro de la humanidad que dependerán del desarrollo científico y tecnológico. Estando de acuerdo con la frase (atribuida a Bohr) "*hacer predicciones es difícil, especialmente sobre el futuro*", no queremos ponernos el gorro de futurólogos y lo apuntado en este capítulo debe considerarse especulativo.

Pero se puede afirmar que no sabemos como será el futuro de la humanidad, no sabemos como será el futuro de la ciencia, ni de la química; lo que sí sabemos es que **SIN CIENCIA NO HAY FUTURO**.

Notas
[1] Realmente no hay alternativa a la medicina, por lo tanto este término es un oximoron.
[2] B. Herradón, Journal of Feelsynapsis 2013, # 11, 46. (JoF 2013, # 11, 46; http://jof.feelsynapsis.com/011/index.html)
[3] G. M. Whitesides, Angew. Chem. Int. Ed. 2004, 43, 3632.
[4] Este aspecto se discute en el artículo The unplanned impact of Mathematics, publicado en la revista Nature 2011, 475, 166.

Bernardo Herradón
Doctor en Química
IQOG - CSIC

Capítulo 73
¿Por qué se forman cristales y por qué lo hacen siempre en determinadas formas?
(Por Miquel Àngel Cuevas-Diarte, Laura Bayés y Teresa Calvet)

Si vamos a hablar de cristales, primero evitemos una confusión bastante habitual. Un cristal no es un vidrio. En castellano a menudo hablamos, incorrectamente, del cristal de la ventana. Deberíamos referirnos al vidrio de la ventana. Un cristal es un sólido donde los átomos, iones o moléculas están ordenados en su interior. Un vidrio es un sólido amorfo, desordenado, obtenido por el enfriamiento rápido de una mezcla de silicatos y otros componentes. Es una forma metaestable. Se trata de un líquido "congelado". Cuando miramos un cristal nos sorprende, de forma inmediata, la existencia de caras planas, formando ángulos precisos, con aristas bien definidas. Sea cual sea el tamaño del cristal. Unas veces las veremos directamente y en otras necesitaremos una lupa, o un microscopio óptico, o incluso un microscopio electrónico. Observando estas morfologías, a veces espectacularmente bellas y perfectas, nace la Cristalografía. De hecho, el termino cristal proviene de la palabra *krystallos* con la que los griegos denominaban al hielo y a aquellos minerales transparentes, como el cristal de roca, que pensaban que se habían formado a partir del frio.

Si un cristal es un sólido ordenado, de este orden se deducen una serie de propiedades. Un cristal es periódico (las unidades que lo componen se distribuyen periódicamente en el espacio), es homogéneo (las propiedades de la materia cristalina son idénticas en cada una de sus partes), es anisótropo (las propiedades de la materia cristalina varían con la dirección, como consecuencia de que la periodicidad en el cristal no tiene porqué ser igual a lo largo de todas las direcciones), y es simétrico (las unidades se distribuyen en el espacio siguiendo unos patrones con un cierto contenido de simetría, que hace equivalentes átomos, iones, o moléculas en el cristal).

¿Y dónde hay cristales? Pues casi en todas partes. Evidentemente es una exageración, pero no tanto. Los cristales están en la base de que podamos utilizar y beneficiarnos de muchas substancias y dispositivos que ayudan a hacernos la vida mucho más agradable. Los cristales y su cristalización juegan un papel muy importante en algunos alimentos como la sal común, el azúcar, la manteca, la leche, los aceites, el chocolate. Son la base del desarrollo y la eficiencia de muchos medicamentos que habitualmente tomamos para el dolor de cabeza, de muelas, etc. y otras enfermedades mucho más complejas. Los pigmentos utilizados desde la antigüedad, y también la cosmética moderna están relacionados con los cristales. Muchos dispositivos electrónicos funcionan gracias a los cristales: tenemos pantallas de televisión y monitores de ordenadores con cristales líquidos, algunos relojes nos dan la hora de forma puntual gracias a alguna propiedad de los cristales, algunos instrumentos médicos y sistemas de transporte de electricidad se basan en materiales superconductores, y muchos otros. Las conchas y caparazones de muchos animales, e incluso las de los huevos son estructuras complejas y muy eficientes parcialmente cristalinas. Nosotros

llevamos cristales en los huesos y en el esmalte de los dientes. Si tenemos la desgracia de necesitar una prótesis, muchas son cristalinas. Mantenemos el equilibrio gracias a un pequeño cristalito situado en la parte interna de nuestro oído. Y sentimos dolores muy agudos si también tenemos la desgracia de sufrir de gota o cristalizamos en nuestro interior cálculos renales o de otro tipo formando cristales que en unos casos martirizan nuestras articulaciones y en otros son demasiado grandes para ser evacuados con facilidad.

¿Y los cristales cómo se forman? Pues vamos a contemplar dos escenarios: en la naturaleza y en el laboratorio. En la naturaleza pueden formarse por evaporación del solvente en el que están disueltos los iones, por ejemplo el agua de una laguna, por enfriamiento de un fluido fundido, por ejemplo durante una erupción volcánica, por efecto de las altas temperaturas y presiones en el interior de la Tierra, etc. Es evidente que no observamos siempre grandes cristales y bien formados, más bien raramente. Lo normal es requerir de instrumentos como la lupa o el microscopio para poderlos ver. Pero el tamaño no importa. Sus características fundamentales son las mismas. También existen cristales grandes, incluso gigantes. En algunas zonas mineras se extraen ejemplares extraordinarios que después podemos admirar en los museos y en algunas tiendas. Juan Manuel García-Ruíz ha descrito magistralmente algunos casos espectaculares en una magnífica película que se puede visualizar por Internet: El Misterio de los Cristales Gigantes. Se trata de los cristales de yeso de Pulpí, en Almeria, de los de la mina de El Teniente en Chile, o de los cristales gigantes de la mina de Naica, en México. En el laboratorio, imitando a la naturaleza, los mecanismos de obtención de cristales también se basan en la evaporación o enfriamiento lento de una solución, en el enfriamiento lento de un fundido, en la sublimación de un sólido y condensación del vapor, en la difusión a través de un medio poroso, y un largo etcétera. Según el material que queramos cristalizar, y según el objetivo que persigamos. Para determinar la estructura interna del cristal cada vez requerimos de cristales de menor tamaño. Y el laboratorio que tenemos siempre más a mano es la cocina de nuestra casa, y en la cocina seguro que tenemos sal común para aderezar nuestros platos. La sal común que utilizamos habitualmente, es un polvo blanco que a simple vista no sabríamos decir si está formado por cristales. Sin embargo, si lo observamos con una lupa, nos aparecerán unos magníficos cristales en forma de cubos, diminutos para que se disuelvan con rapidez, pero con una morfología bien desarrollada. La sal común, químicamente cloruro sódico, cristaliza en el sistema cúbico, y da lugar a cubos con los tres lados iguales y los ángulos de 90° entre las tres direcciones fundamentales. Esos cubos son el reflejo del ordenamiento exacto de los iones de sodio y de cloro, alternados para compensar sus cargas eléctricas, en una red cúbica limitada por las caras. La sal común puede proceder de una mina de sal o, más habitualmente, de la evaporación del agua de mar en salinas que todos hemos visto alguna vez.

Hagamos un ejercicio. Ya que estamos en la cocina, cojamos un recipiente y vertamos en él una o dos tazas pequeñas de agua del grifo. Ahora calentemos esa agua en el fuego y echemos un poco de sal removiendo para favorecer su disolución. Al principio, la sal se disuelve rápidamente. Si continuamos añadiendo sal al agua, cada vez se disolverá más lentamente a pesar de que la temperatura aumenta su solubilidad. Cuando veamos que ya no se disuelve más, paramos. Hemos obtenido una solución saturada de cloruro sódico en agua. Cada substancia tiene lo que denominamos una curva de solubilidad que determina la cantidad que puede disolverse a una temperatura dada. Podíamos haber buscado la curva de solubilidad del cloruro sódico en Internet y hubiéramos sabido la cantidad de sal a utilizar para una cantidad determinada de agua a

la temperatura que mediríamos con un termómetro. Nosotros hemos trabajado empíricamente, pero llegamos a un resultado final suficiente para nuestros propósitos. Ahora retiremos el recipiente del fuego, pongámoslo en un lugar resguardado de cambios bruscos de temperatura, tapémoslo mínimamente para que no caiga polvo y dejémoslo en reposo. Al cabo de horas o días, según la cantidad utilizada, veremos que se han formado cubos de sal común en mayor o menor cantidad (figura 1). No es magia. Es pura Cristalografía. La solución, con el tiempo, se ha enfriado y ha ido evaporando el solvente. Se ha sobresaturado. Ha llegado a una concentración en la que se han formado primero núcleos de sal invisibles a simple vista y esos núcleos han ido creciendo paulatinamente. Se trata del fenómeno de nucleación y crecimiento en el que se basa el crecimiento cristalino. Observarlo en una lupa o en un microscopio es una maravilla y una bella sorpresa para la mayoría. Evidentemente, no vamos a estar días pegados al microscopio, pero podemos grabarlo y observarlo posteriormente. O podemos trabajar con una sola gota con lo que el proceso será rápido aunque se formaran muy pocos cristales.

Pero si en nuestra cocina tenemos también una sal que se denomina sal Maldon y la observamos con la lupa, no veremos cubos. Veremos pirámides. Muy bellas. Con las líneas de crecimiento muy bien marcadas y huecas por dentro al haber crecido muy rápidamente las aristas. Pero habíamos dicho que el cloruro sódico es cúbico. Efectivamente, pero ello no significa que los cristales cúbicos cristalicen siempre en forma de cubos. También pueden desarrollarse caras de octaedro, de tetraedro, etc. Cada substancia tiende a desarrollar una morfología de equilibrio en función de su estructura interna que, sin entrar en detalles depende de las direcciones de enlaces fuertes, que en el caso del cloruro sódico es el cubo. Pero variando las condiciones experimentales de crecimiento o añadiendo cantidades y tipos de impurezas podemos variar la morfología del cristal parcial o totalmente.

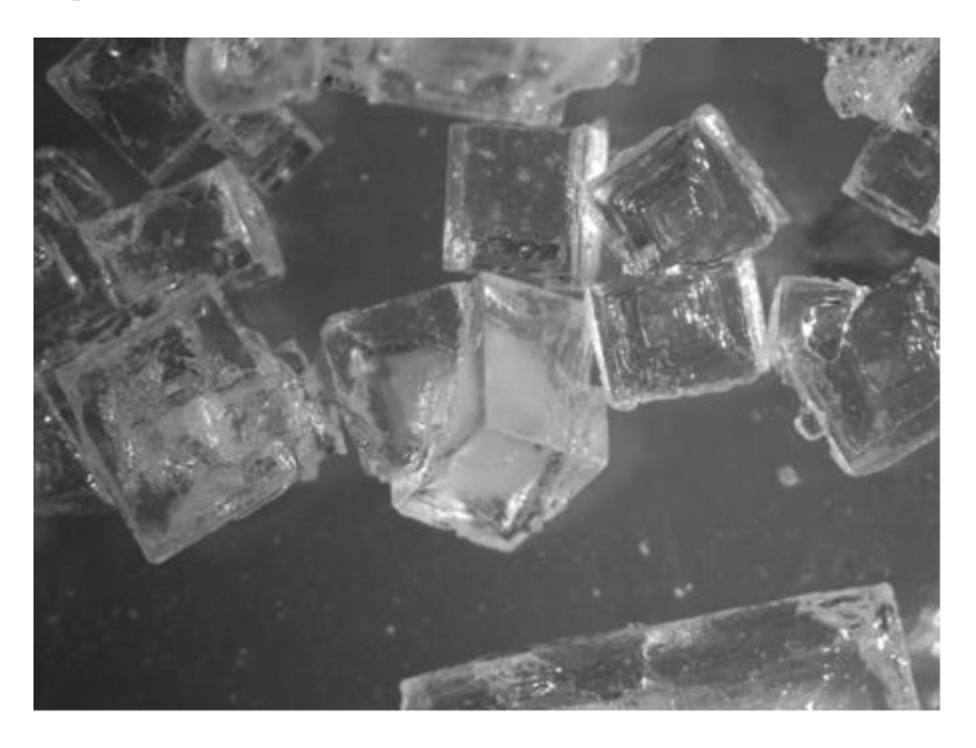


Figura 1. *Cristales de Cloruro Sódico (sal común) observados con una lupa. Se ve perfectamente la simetría cúbica. En algunos de ellos también quedan marcadas las líneas de crecimiento.*

Por ejemplo, con la sal común si utilizamos formamida en lugar de agua, obtendremos octaedros de sal. Si hemos hablado de las salinas, igual recordamos que no siempre se ven las balsas de evaporación de color blanco. Algunas de ellas las veremos coloreadas y de colores diferentes según la zona geográfica. Ello es debido a la presencia de diferentes sales presentes en el agua. De forma análoga, las industrias de la sal trabajan constantemente para incorporar otros elementos que puedan proporcionarnos algún efecto beneficioso para nuestra salud o placentero para nuestro paladar. Es cierto que no deberíamos abusar de la sal, sobre todo algunos de nosotros propensos a algunas enfermedades. Por ello se trabaja también para reducir la cuota de sal que tomamos sin disminuir la sensación de que estamos tomando sal. El incremento de la superficie específica de los cristales de sal va en esta dirección y ello depende en gran medida de los métodos de cristalización que bien controlados y basados en conceptos científicos pueden favorecer la formación de láminas, pirámides, esferas huecas, etc.

El estudio o la descripción de un cristal lo podemos abordar a dos niveles: macroscópico y microscópico. Los dos estrechamente relacionados. El orden de los átomos, iones o moléculas en el interior del cristal es la causa de que los cristales formen poliedros regulares más o menos bien formados a escala macroscópica. Dicho de otra manera, las formas externas de los cristales son el efecto de su orden interno. La simetría macroscópica es una simetría finita, ya que los poliedros cristalinos son figuras finitas, y el número de operaciones de simetría que los describen también es finito. En cambio, la simetría microscópica es infinita ya que la distribución de átomos, iones o moléculas en el interior del cristal se puede considerar infinita, aunque limitada por las caras. El interior del cristal se describe a la escala del Angstrom. Recordemos que un metro son 10^{10} Å, es decir, diez mil millones de Angstroms. Inicialmente, y hasta el año 1912 en el que se descubre la difracción de rayos X, que abre las puertas a la posibilidad de determinar la estructura interna de los cristales, a partir de su observación macroscópica y de algunas pocas propiedades mesurables con instrumentos rudimentarios tan solo se podían establecer hipótesis sobre su interior, aunque muchas de ellas fueron corroboradas posteriormente. Fue necesario reinventar la noción de átomo, y después introducir la de molécula, para poder comenzar a introducir la noción de que en el interior del cristal hay una pequeña parte que se repite periódicamente en las tres direcciones del espacio y que nos describe el conjunto del cristal. La estructura interna del cloruro sódico nos muestra el ordenamiento preciso de los iones de sodio y de cloro en las tres direcciones del espacio que adoptan una simetría cúbica, con vectores de periodicidad de módulos idénticos en las tres direcciones y formando ángulos rectos entre ellos (figura 2). Este ordenamiento interno se traduce en los cubos de sal común que observamos macroscópicamente. Una substancia cristalina tiene una estructura interna determinada y esta se traduce en una forma externa acorde con ella.

Este orden interno, microscópico, comporta un cierto contenido de simetría. Aquí hablaremos de simetría microscópica, de simetría espacial, de simetría infinita. Dada la escala a la que nos situamos, el cristal puede considerarse como un universo infinito, limitado por las caras. Las operaciones de simetría hacen equivalentes átomos, iones o moléculas. Son las mismas operaciones que hemos descrito en la simetría macroscópica, y algunas más: la translación, que asociada a la reflexión da lugar a los planos de deslizamiento, y asociada a la rotación a los ejes helicoidales.

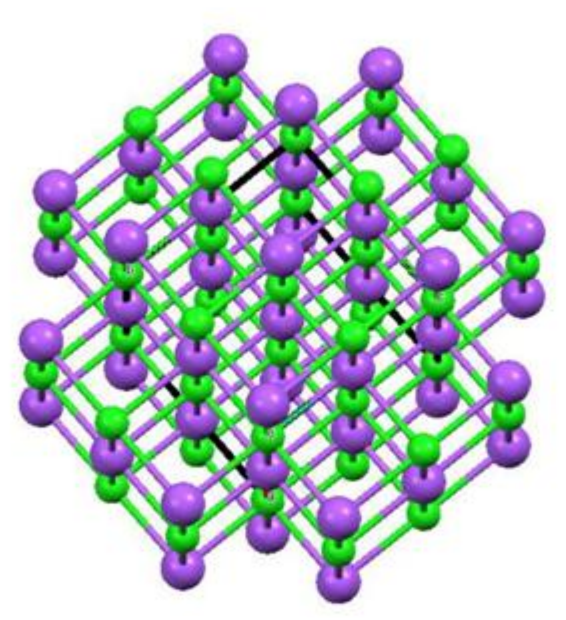


Figura 2. *Estructura cristalina del cloruro sódico (sal común). Las esferas violetas representan iones de cloro y los verdes iones de sodio.*

También aquí el conjunto de elementos de simetría que describen el interior de un cristal forman un grupo de operaciones en el sentido matemático del término. En dos dimensiones son 17. Es la simetría presente en cualquier dibujo u ornamentación que pretenda llenar todo el plano a partir de un motivo repetido periódicamente por dos translaciones (la mayor parte de papeles pintados que decoran las paredes de las habitaciones, la ornamentación de las paredes y pavimentos en la arquitectura mudéjar de Aragón o de la Alhambra de Granada, o de Los Reales Alcázares de Sevilla, los pavimentos de muchas aceras de nuestras ciudades. En tres dimensiones son 230 los grupos de operaciones de simetría espacial o de simetría infinita. Si introducimos diferencias de color (simetría poli cromática) los grupos aumentan a 1651 en simetría dicromática. Cada uno de estos grupos de operaciones de simetría espacial es el resultado de combinar una red de puntos, que representa la distribución periódica de los átomos, iones o moléculas en el interior del cristal, con el conjunto de operaciones de simetría que representan las equivalencias entre los átomos, iones o moléculas en el espacio. En dos dimensiones consideramos cinco redes planas en función de la simetría que pueden contener, y en tres dimensiones son 14. En todos los casos la geometría de la red queda configurada por la simetría que puede contener.

La forma de los cristales depende de la velocidad relativa de crecimiento de las diferentes caras. Aquellas que tienen una velocidad de crecimiento más baja serán las que tenderán a predominar en la forma cristalina. Las de velocidades superiores tienden a desaparecer. Las velocidades de crecimiento de las caras dependen de la distribución espacial de las unidades de crecimiento en el interior del cristal. Esta distribución es una consecuencia de las fuerzas de atracción y de la distribución de las fuerzas de enlace entre las unidades estructurales en cada cara. La forma externa del cristal no deja de ser una buena indicación de la distribución simétrica de las unidades estructurales en su interior.

Cuando observamos un cristal que durante el proceso de crecimiento haya desarrollado sus caras suficientemente, podemos apreciar que algunas de estas caras son

exactamente idénticas las unas a las otras después de efectuar algunos movimientos en el espacio. Es lo que denominamos una operación de simetría. Cada poliedro tiene unos elementos de simetría característicos que lo describen y que permiten su clasificación. El conjunto de elementos de simetría de un poliedro forman un grupo de operaciones en el sentido matemático del término. Como todos los elementos de simetría de un poliedro cristalino se intersectan en un punto (su centro), hablamos de grupo puntual de simetría. Las operaciones de simetría son pocas: rotación, reflexión, inversión, y rotación seguida de inversión. Todas estas operaciones de simetría, que aquí aplicamos al cristal, no dejan de ser las mismas que podemos aplicar a los objetos que nos rodean. Por ejemplo, una mesa de comedor rectangular presenta dos planos de reflexión perpendiculares entre sí, i un eje de rotación de 180° en la intersección de los planos. Si la mesa es cuadrada, el ángulo de rotación del eje será de 90°. Muchos edificios los podemos dividir en dos mitades exactamente idénticas por un plano de reflexión.

¿Y por qué los cristales adoptan siempre determinadas formas? Porque la forma externa de un cristal refleja la simetría del grupo puntual. Los cristales están formados por una o varias formas cristalinas, todas ellas correspondientes al grupo puntual del cristal. Una forma cristalina es el conjunto de caras relacionadas entre ellas por la simetría del grupo puntual. La estructura interna del cristal induce un hábito característico de la forma externa cristalina. Las unidades estructurales del cristal pueden estar distribuidas de forma similar en las tres direcciones del espacio (habito isométrico), estar distribuidas de forma bidimensional (hábito laminar), tener una dimensión más desarrollada que las otras dos (hábito prismático), o estar distribuidas prácticamente según una dimensión (hábito acicular), como puede verse en la Figura 3.

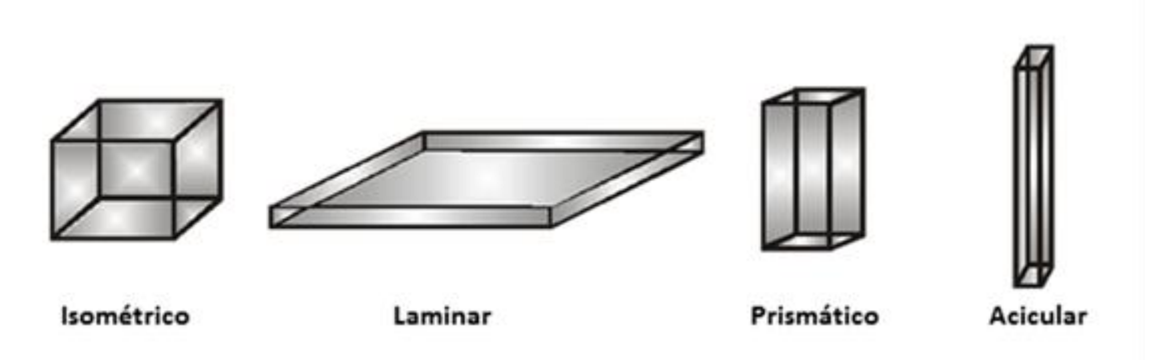


Figura 3: *Diferentes hábitos cristalinos.*

Además, durante el proceso de crecimiento (en la naturaleza o en el laboratorio), los cristales pueden dar lugar a maclas, que son asociaciones de cristales siguiendo una cierta ley de simetría. Como, por ejemplo, la macla de contacto de la albita, la macla de rotación de la ortoclasa, la macla de interpenetración de la fluorita, la macla cíclica del aragonito, etc.

Miquel Àngel Cuevas-Diarte **Laura Bayés** **Teresa Calvet**
Doctores en Cristalografía
Profesores de la Universitat de Barcelona

Capítulo 74
¿Qué son los cuasicristales?
(Por Carlos M. Pina)

Buena parte de la materia sólida que vemos a nuestro alrededor se encuentra en lo que se conoce como estado cristalino. Este estado consiste en una ordenación periódica de millones de átomos a lo largo de las tres direcciones del espacio. Dependiendo de la fórmula química de cada compuesto y de las condiciones de presión y temperatura, el ordenamiento atómico de los cristales podrá ser diferente, pero éste siempre será periódico. A finales del siglo XIX, los cristalógrafos demostraron que solo existen 230 esquemas estructurales diferentes para ordenar átomos de forma periódica en el espacio tridimensional. Estos esquemas se denominan grupos espaciales.

A cualquier sustancia que cristalice a una presión y temperatura determinadas se le puede asignar uno y solo uno de estos 230 grupos espaciales. La limitación de los esquemas estructurales de los cristales se debe, precisamente, a la naturaleza periódica de la disposición de sus átomos. Aunque la deducción de los 230 grupos espaciales tridimensionales es compleja e implica elaborados razonamientos geométricos, podemos adquirir una buena idea de su fundamento con un simple ejemplo limitado a las dos dimensiones. Imaginemos que queremos rellenar completamente una superficie con polígonos iguales de cartón colocándolos unos junto a otros. Tarde o temprano nos daremos cuenta de que eso solo es posible si lo hacemos con triángulos equiláteros, rombos, romboides, rectángulos, cuadrados o hexágonos. Si lo intentamos hacer con otros polígonos como, por ejemplo, pentágonos o decágonos veremos que no hay forma de colocarlos sin que queden huecos entre ellos (ver figura 1).

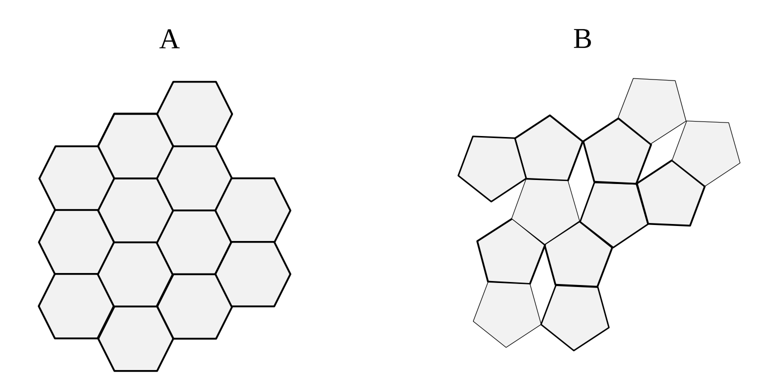


Figura 1: *A) Relleno completo de una superficie con hexágonos. B) Imposibilidad de rellenar completamente una superficie con pentágonos.*

En el caso tridimensional podríamos hacer algo similar a lo que hemos hecho en dos dimensiones y comprobaríamos que seríamos capaces de rellenar completamente y de forma periódica el espacio con ciertos poliedros (como cubos, prismas de base cuadrada, prismas hexagonales, etc.) Sin embargo, no podríamos hacerlo con otros poliedros como, por ejemplo, dodecaedros o icosaedros. Esto quiere decir que la formación de un cristal periódico solo es posible mediante la colocación ordenada de unidades estructurales con formas específicas. Estas unidades estructurales se denominan celdas cristalinas elementales y contienen los átomos que forman cada compuesto químico.

Una consecuencia del ordenamiento periódico de las celdas elementales en el interior de los cristales es que éstos muestran formas poliédricas externas con simetrías compatibles con dicho ordenamiento. Así, dentro de la inmensa variedad de morfologías que presentan los cristales en la naturaleza encontraremos cubos, prismas, pirámides y combinaciones de estas formas, pero nunca veremos dodecaedros e icosaedros regulares, ni tampoco poliedros que tengan simetrías de orden cinco. Desde el punto de vista de la Cristalografía clásica, estas simetrías de orden cinco se consideran "prohibidas" por ser incompatibles con la periodicidad interna de las estructuras de los cristales.

La naturaleza periódica de las estructuras cristalinas se demostró gracias a los experimentos de difracción de rayos X realizados por Max von Laue y colaboradores en 1912 [1]. El experimento original de Laue consistió en hacer incidir un haz de rayos X sobre un cristal, detrás del cual se había colocado una placa fotográfica. Los rayos X, al atravesar el cristal, interfirieron con sus átomos y produjeron un patrón de difracción que quedó impresionado en la placa fotográfica. Este patrón resultó ser una distribución periódica de puntos (máximos de difracción), algo que solo puede ocurrir cuando el objeto que produce la difracción es también periódico. En las décadas siguientes al experimento de Laue se registraron miles de patrones de difracción de cristales, tanto empleando rayos X como haces de electrones acelerados. Todos esos patrones resultaron ser patrones periódicos y su análisis permitió determinar y estudiar las estructuras cristalinas de innumerables compuestos, tanto naturales como sintéticos. La Cristalografía había confirmado experimentalmente su hipótesis de la periodicidad de las estructuras cristalinas y contaba con una herramienta formidable para investigarlas: la difracción.

Pero en 1982 sucedió algo sorprendente. En el transcurso de una investigación sobre ciertas aleaciones sintéticas de aluminio y manganeso, el científico Dan Shechtman observó unos enigmáticos patrones de difracción de electrones que tenían simetría pentagonal [2]. Esos patrones de difracción anómalos no podían corresponder a un ordenamiento periódico de los átomos de las aleaciones que estaba estudiando, pues mostraban simetrías cristalográficamente prohibidas. Sin embargo, estaba claro que la difracción había sido producida por una estructura muy ordenada. ¿Era posible que existiera un estado de la materia en el que los átomos estuvieran ordenados de forma diferente a como lo hacen dentro de los cristales? La publicación de los patrones de difracción de Shechtman desconcertó a los cristalógrafos pues constituía un desafío a los fundamentos de la Cristalografía. Durante varios años, los científicos, con el prestigioso e influyente premio Nobel de Química Linus Pauling a la cabeza, intentaron conciliar los principios de la Cristalografía clásica con las observaciones de patrones de difracción con simetrías "no cristalográficas". Para ello idearon complejos modelos estructurales según los cuales la asociación de un gran número de cristales periódicos

con diferentes orientaciones podría generar patrones de difracción con simetrías de orden cinco o diez. Estos modelos resultaron ser excesivamente complicados y muchas veces forzados y, finalmente, la solución al enigma de los patrones de Shechtman resultó ser tan sencilla como extraña: los átomos de algunos materiales pueden ordenarse en su interior a lo largo de grandes distancias de manera no periódica. Para entender ese nuevo tipo de ordenamiento atómico resultó fundamental el trabajo del matemático Roger Penrose sobre teselados no periódicos [3]. Penrose estableció una serie de reglas de ordenamiento de paralelogramos que permitían rellenar el espacio bidimensional de forma completa pero sin que ese ordenamiento fuera periódico (ver figura 2).

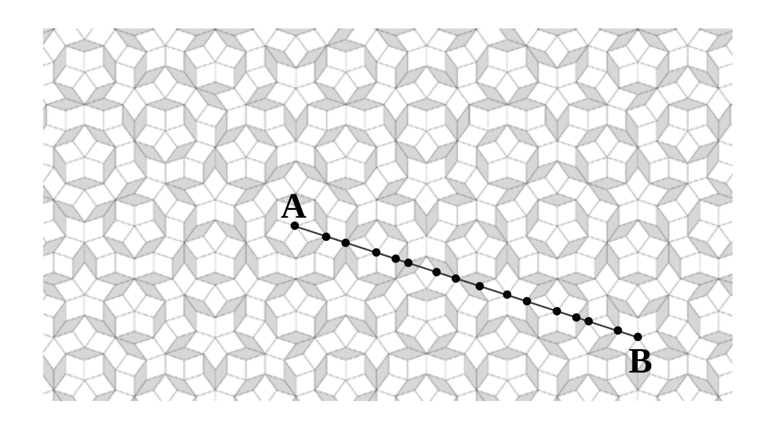
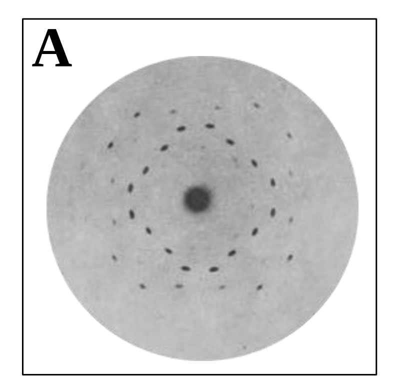


Figura 2. *Teselado de Penrose en el que la superficie queda completamente cubierta al combinar rombos de dos tamaños según ciertas reglas de construcción. Como puede verse, la simetría de este teselado es localmente pentagonal en los puntos A y B pero la secuencia de los vértices de los rombos entre esos dos puntos no es periódica.*

En tres dimensiones se pueden establecer unas reglas de ordenamiento de poliedros, similares a las que generan los teselados de Penrose, que permitan el relleno no periódico del espacio tridimensional. Cualquier compuesto químico cuyos átomos se sitúen en los vértices de los poliedros dispuestos según esas reglas tendrá una estructura cuasiperiódica y producirá patrones de difracción con simetrías "no cristalográficas". Uno de esos compuestos fue precisamente el que encontró Shechtman cuando investigaba sus aleaciones de aluminio y manganeso. Los materiales con este tipo de ordenamiento interno de sus átomos se denominan desde entonces *cuasicristales*, un nombre que proviene de la contracción de "cristales cuasiperiódicos".

El descubrimiento de los cuasicristales obligó a los científicos a revisar los conceptos cristalográficos de ordenamiento atómico y estructura cristalina. Una consecuencia de esa revisión del paradigma cristalográfico fue la redefinición de cristal. En 1991, la International Union of Crystallography estableció que *un cristal es todo aquel sólido que tiene un diagrama de difracción esencialmente discreto* (1). Como puede verse, en esta definición se ha excluido toda referencia al orden periódico y la naturaleza cristalina (o cuasicristalina) de un material queda evidenciada por la difracción que produce, independientemente del tipo de orden que muestran sus átomos (ver figura 3).



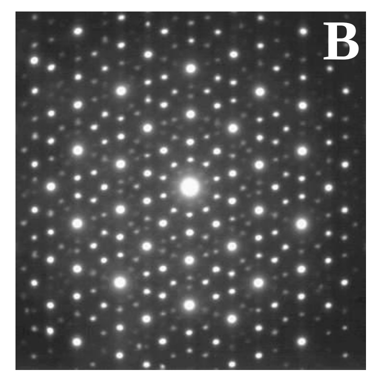


Figura 3. *Diagramas de difracción producidos por un cristal y un cuasicristal.* ***A)*** *Diagrama de difracción de rayos X con simetría cuaternaria correspondiente a un cristal cúbico de sulfuro de zinc (blenda) [4].* ***B)*** *Diagrama de difracción de electrones de un cuasicristal icosaédrico de la aleación HoMgZn (2). Nótese la simetría de orden diez de la distribución de puntos de difracción.*

Los cuasicristales poseen algunas propiedades específicas relacionadas con la aperiodicidad de los enlaces atómicos dentro de sus estructuras: son relativamente frágiles, malos conductores de la electricidad y del calor, y muestran bajos coeficientes de fricción. Aprovechando estas propiedades, algunos materiales cuasicristalinos se están empezando a utilizar como aislantes térmicos, como componentes de algunos LEDs, en sistemas que convierten calor en electricidad, e incluso se han fabricado sartenes antiadherentes con ellos.

Aunque la gran mayoría de los cuasicristales que se conocen en la actualidad han sido sintetizados en el laboratorio, recientemente se han descubierto los primeros cuasicristales naturales. El hallazgo se produjo al estudiar un meteorito encontrado en 1979 en la región de Khatyrka en la Península de Kamchatka (Rusia) y que había sido adquirido en 1990 por el Museo de Historia Natural de Florencia (Italia).

El equipo del profesor Luca Bindi ha logrado identificar mediante difracción de electrones dos cuasicristales diferentes dentro de ese meteorito: uno con simetría icosaédrica y denominado *icosaedrita* por la *International Mineralogical Asociation* y otro, con simetría decagonal, llamado *decagonita* [5,6]. Según recientes experimentos de síntesis llevados a cabo empleando elevadísimas presiones, la formación de cuasicristales como los descubiertos en el meteorito de Khatyrka se debería a la colisión de asteroides dentro de nuestro sistema solar [7]. Sin embargo, es posible que los cuasicristales puedan formarse también bajo otras condiciones y no está completamente descartado que se encuentren en el futuro minerales cuasicristalinos de origen terrestre.

Cuando Isaac Asimov publicó su libro “Cien preguntas básicas sobre la Ciencia” faltaba casi una década para que se descubrieran los cuasicristales y algunos años más para que la comunidad científica aceptara que eran materiales con una ordenación

atómica diferente de la que presentan los cristales. Teniendo en cuenta lo que sabemos actualmente sobre los cuasicristales, y lo que nos queda aún por saber, seguramente Asimov no se resistiría hoy a incluir un capítulo sobre ellos en una edición revisada de su libro.

Notas:
(1) http://reference.iucr.org/dictionary/Crystal.
(2) https://commons.wikimedia.org/w/index.php?curid=10094837.

Bibliografía:
[1] Interferenz-Erscheinungen bei Röntgenstrahlen (1912) W. Friedrich, P. Knipping & M. Laue, München : Verl. der Königlich Bayer. Akad. der Wiss., S. 303 – 322, 363 – 373.
[2] Metallic phase with long-range orientational order and no translational symmetry (1984) D.S. Shechtman, I. Blech, D. Gratia. J.W.Cahn. Physical Review Letters 53 (20): 1951. 2.
[3] The role of Aesthetics in Pure and Applied Mathematical Research (1974) R. Penrose. Bulletin of the Institute of Mathematics and its Applications. 10: 266Interferenzerscheinungen bei Röntgenstrahlen (1913) W. Friedrich, P. Knipping, M. Laue, M. Annalen der Physik 346 (10): 971.
[4] Interferenzerscheinungen bei Röntgenstrahlen (1913) W. Friedrich, P. Knipping, M. Laue, M. Annalen der Physik 346 (10): 971.
[5] Icosahedrite, $Al_{63}Cu_{24}Fe_{13}$, the first natural quasicrystal (2011) L. Bindi, P.J. Steinhardt, N. Yao, P.J. Lu. American Mineralogist 96 (5-6): 928.
[6] Natural quasicrystal with decagonal Symmetry (2015) L. Bindi, N. Yao, Ch. Lin, L.S. Hollister, Ch.L Andronicos, V.V. Distler, M.P. Eddy, A. Kostin, V. Kryachko, G.J. MacPherson, W.M. Steinhardt, M. Yudovskaya P.J. Steinhardt. Scientific Reports 5:9111, PMID 25765857.
[7] Shock synthesis of quasicrystals with implications for their origin in asteroid collisions. (2016) P. D. Asimow, Ch. Lin, L. Bindi, Ch. Maa, O. Tschaunere, L. S. Hollisterg, P. J. Steinhardt Proceedings of the National Academy of Sciences of the United States of America 113(26):7077

Carlos M. Pina
Doctor en Ciencias Geológicas
Profesor Titular, Universidad Complutense de Madrid

Capítulo 75
¿Qué es el hidrógeno metálico? ¿Cómo puede ser el hidrógeno un metal?
(Por Carlos M. Pina)

La mayoría de los elementos químicos de la tabla periódica se clasifican como metales. Los metales poseen una serie de propiedades comunes, entre las cuales destacan su brillo característico, su alta conductividad eléctrica y térmica, su maleabilidad y su ductilidad. Todas estas propiedades se pueden explicar teniendo en cuenta el tipo de enlace que une a sus átomos. En los metales, los núcleos atómicos se encuentran muy próximos unos a otros, rellenando el espacio de la forma más eficiente posible y permaneciendo unidos entre sí gracias a una nube o gas de electrones que los envuelven. Los electrones de esa nube, además de reflejar la luz y proporcionar a los metales su especial brillo, tienen una gran libertad de movimiento, lo que explica que la electricidad y el calor se transmitan fácilmente a través de ellos. Además, la relativa debilidad de los enlaces que forman los metales permite que sus átomos puedan deslizarse unos respecto a otros. Ello tiene como consecuencia que se puedan deformar y estirar con facilidad. Un caso extremo de deslizamiento de átomos en un metal lo encontramos en el mercurio, que a temperatura ambiente no presenta una estructura cristalina como la mayoría de los metales sino que es líquido.

Como bien nos dice Isaac Asimov en su libro "Cien preguntas básicas sobre la Ciencia", para que se produzca el enlace metálico y existan electrones móviles es necesario que entre el núcleo atómico y los electrones más externos que se encuentran a su alrededor haya un número de capas electrónicas que apantallen la atracción electrostática que ejercen los núcleos (con carga positiva) sobre los electrones (con carga negativa). Estas capas son más numerosas en los elementos químicos con un alto número atómico y que, por lo tanto, poseen un gran número de electrones. Este es el caso, por ejemplo, del potasio, el hierro o el oro. A diferencia de estos átomos, el átomo de hidrógeno solo tiene un protón y un electrón, que suele compartir con el electrón de otro átomo de hidrógeno para formar la molécula gaseosa H_2. Al no haber apantallamiento posible, pues no existen capas electrónicas inferiores, los electrones compartidos están fuertemente ligados a sus núcleos en la molécula de H_2 y, en consecuencia, el hidrógeno no tiene propiedades metálicas en condiciones ambientales. Sin embargo, se ha especulado mucho sobre la posibilidad de que el hidrógeno se convierta en un metal bajo condiciones de elevada presión y temperatura.

En 1935, los físicos Eugene Wigner y Hillard Bell Huntington predijeron que para que el hidrógeno se transforme en un metal serían necesarias elevadísimas presiones [1]. Se estima que una presión superior a la que existe en el núcleo terrestre (unos 3,5 millones de atmósferas) podría obligar a los protones del hidrógeno a empaquetarse de forma compacta. Solo entonces los electrones quedarían libres y se obtendría hidrógeno metálico líquido, algo parecido al mercurio. No obstante, el hidrógeno metálico sería un compuesto bastante diferente de los metales que conocemos

y se presentaría como un estado de la materia degenerado con propiedades singulares. Así, se piensa que el hidrógeno metálico podría ser superconductor a temperatura ambiente y comportarse también como un superfluido. Por otro lado, algunos científicos creen que el hidrógeno metálico podría permanecer de forma metaestable durante cierto tiempo, es decir sin transformarse inmediatamente en hidrógeno ordinario una vez eliminada la presión necesaria para formarlo. Si esto fuera así, el hidrógeno metálico líquido podría emplearse como un combustible limpio (pues su combustión solo produciría agua) y con una eficacia energética casi cinco veces superior a los combustibles H_2/O_2 empleados actualmente. El empleo de este nuevo combustible permitiría, entre otras cosas, aumentar la potencia de las naves espaciales y, por tanto, acortar considerablemente la duración de los viajes a otros planetas. Este potencial uso del hidrógeno metálico como combustible, junto con sus propiedades electrónicas y sus posibles aplicaciones tecnológicas derivadas, han hecho de su obtención una importante meta científica.

A pesar del gran interés que existe por el hidrógeno metálico y del gran número de experimentos realizados para obtenerlo, los resultados han sido hasta la fecha limitados. En 1996, un grupo de científicos del *Lawrence Livermore National Laboratory* (EE.UU.) comunicó que había conseguido casualmente detectar durante un milisegundo la formación de hidrógeno metálico durante un experimento en el que sometieron hidrógeno molecular líquido a temperaturas de varios miles de grados y presiones de algo más de un millón de atmósferas [2]. Este resultado fue en cierto modo sorprendente, pues experimentos previos realizados por otros investigadores empleando hidrógeno molecular sólido y presiones de hasta 2,5 millones de atmósferas no habían dado lugar a la formación de hidrógeno metálico.

Desde entonces, los intentos para obtener hidrógeno metálico han continuado en laboratorios de todo el mundo. En 2011, científicos del *Max Planck Institut* (Alemania) publicaron un artículo en el que afirmaron haber conseguido hidrógeno metálico a presiones entre 2,6 y 3 millones de atmósferas, pero su hallazgo fue posteriormente cuestionado por otros investigadores [3,4]. En 2015, un grupo de investigadores de los *Sandia National Laboratories* (EE.UU) publicaron unos prometedores resultados tras llevar a cabo una serie de experimentos empleando la llamada *máquina Z*, un moderno generador de ondas electromagnéticas de alta frecuencia [5]. Estos experimentos se realizaron aplicando ondas de choque combinadas con enormes campos magnéticos y parece que han constituido un claro progreso en una búsqueda que dura ya varias décadas. Sin embargo, y a pesar de los avances realizados en los últimos años, las condiciones de presión y temperatura para la síntesis del hidrógeno metálico todavía no están totalmente definidas y la investigación prosigue en la actualidad.

La búsqueda del hasta ahora esquivo hidrógeno metálico no se limita a los experimentos de laboratorio, sino que se extiende al espacio. Desde hace tiempo, los científicos piensan que existen grandes masas de hidrógeno metálico líquido en el interior de Júpiter y quizá también en algunos grandes planetas extrasolares. Júpiter es el mayor de los planetas de nuestro sistema solar y su masa es unas 320 veces mayor que la de la Tierra. Su atmósfera está formada por 90 % de hidrógeno, 10% de helio y una cantidad inferior al 0.1 % de metano, agua, amoníaco, sulfuro de hidrógeno y otros gases. Las capas interiores del planeta están también mayoritariamente compuestas por hidrógeno y se cree que su relativamente pequeño núcleo es rocoso. Debido a su enorme masa, la presión en Júpiter aumenta desde unas dos atmósferas en zonas de su superficie hasta unos 100 millones de atmósferas en su núcleo. Los científicos están convencidos

de que el aumento de presión en el interior de Júpiter tiene que resultar inevitablemente en la transformación del hidrógeno molecular en hidrógeno metálico a una cierta profundidad, si bien todavía no se sabe cuál es esa profundidad. La confirmación de la existencia de una capa de hidrógeno metálico y la medida de su espesor resultarán fundamentales para explicar cómo se genera el enorme campo magnético de Júpiter. Actualmente se piensa que este campo magnético se debe a la combinación de grandes masas de hidrógeno metálico (que se comporta como un excelente conductor de electrones) con la rápida rotación de Júpiter, cuyo día apenas dura 10 horas.

El pasado 4 de julio de 2016 llegó a Júpiter, después de casi cinco años de viaje, la sonda espacial *Juno*. Enviada por la NASA, *Juno* tiene como misión principal estudiar la gravedad y los campos magnéticos de Júpiter. Para ello orbitará alrededor de los polos del planeta hasta febrero de 2018. Durante las 37 órbitas que realizará enviará constantemente a la Tierra medidas del campo magnético y de la composición de Júpiter (1). El análisis de esas medidas permitirá comprender mejor cómo funciona la inmensa dinamo de Júpiter y qué papel juega el hidrógeno metálico en ella. Mientras tanto, los científicos seguirán intentando en la Tierra sintetizar esa extraordinaria forma de la materia.

Notas:
(1) Para más información sobre la sonda espacial *Juno* el lector puede consultar la siguiente página web: https://www.nasa.gov/mission_pages/juno/main/index.html

Bibliografía:
[1] On the possibility of a metallic modification of hydrogen (1935) E. Wigner & H.B. Huntington. Journal of Chemical Physics 3 (12): 764.
[2] Metallization of fluid molecular hydrogen at 140 GPa (1.4 Mbar) (1996) S.T. Weir, A.C. Mitchell & W. J. Nellis. Physical Review Letters 76: 1860.
[3] Conductive dense hydrogen (2011) M.I. Eremets & I.A. Troyan. Nature Materials 10, 927–931.
[4] Has Metallic Hydrogen Been Made in a Diamond Anvil Cell? (2012) Nellis, W.J., Arthur L. Ruoff & Isaac F. Silvera. arXiv:1201.0407.
[5] Direct observation of an abrupt insulator-to-metal transition in dense liquid deuterium (2015) M. D. Knudson1, M. P. Desjarlais, A. Becker, R. W. Lemke, K. R. Cochrane, M. E. Savage, D. E. Bliss, T. R. Mattsson & R. Redmer. Science 348 (6242):1455.

Carlos M. Pina
Doctor en Ciencias Geológicas
Profesor Titular, Universidad Complutense de Madrid

Capítulo 76
¿Qué son las vitaminas y por qué las necesitamos?

(Por Mayte Conejero Muriel)

76.1 ¿Qué son las vitaminas?

El término vitamina, cuyo origen proviene del latín *vita* ¨vida¨ y del sufijo *amina*, es definido por la Real Academia Española de la Lengua como una ¨*sustancia orgánica que se encuentra en alimentos y que, en cantidades pequeñas, es esencial para el desarrollo del metabolismo de los seres vivos; el organismo no puede fabricar esta sustancia por sí mismo*¨. Por tanto, las vitaminas son imprescindibles para el correcto funcionamiento y desarrollo de los seres humanos. Sin embargo, ¿qué hay de cierto cuando te aconsejan tomar muchas vitaminas o complejos vitamínicos para crecer más, sentirte mejor, más animado y con más energía? ¿Es bueno ingerir cuantas más vitaminas mejor? ¿Es aconsejable tomar complejos vitamínicos comerciales además de nuestra dieta diaria? Como veremos a lo largo de este capítulo, en este caso no es del todo cierto el ¨cuanto más, mejor¨, e incluso en algunos casos puede llegar a ser perjudicial para la salud.

76.2 Las vitaminas y su clasificación.

En primer lugar, debemos saber que son 13 las vitaminas que podemos encontrar en el cuerpo humano (cuyas estructuras químicas se resumen en la Figura. 1), las cuales se clasifican en dos grandes grupos: las vitaminas hidrosolubles (solubles en agua) y las vitaminas liposolubles (solubles en lípidos).

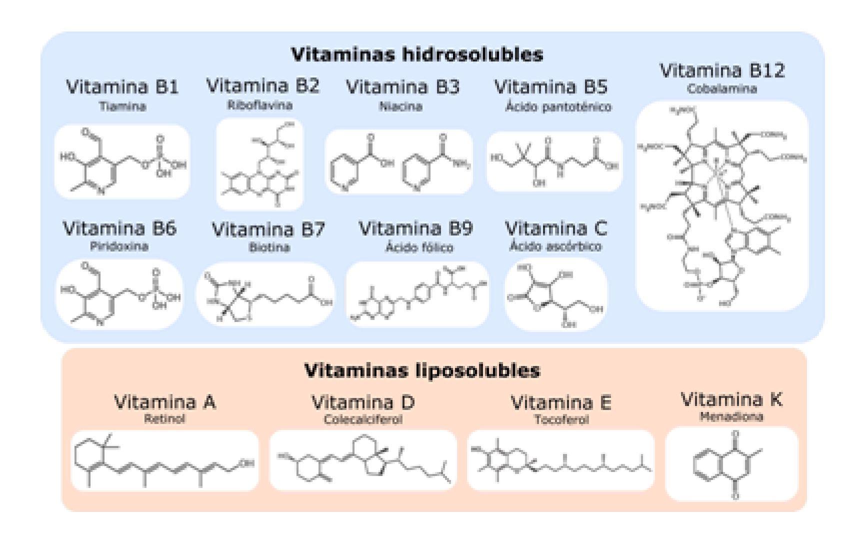


Figura 1: *Clasificación y estructura química de las vitaminas.*

Puesto que no podemos producirlas directamente, existe una cantidad recomendable de ingesta para cada una de ellas en función de sus características y del grupo al que pertenezcan. Y debemos ser conscientes de que la falta o defecto de vitaminas en nuestro cuerpo (hipovitaminosis) puede hacer que éste no funcione adecuadamente y provocarnos enfermedades; sin embargo, no podemos obviar que la ingesta de vitaminas en exceso (hipervitaminosis) también puede resultar tóxica y perjudicial para nuestra salud.

El grupo de las vitaminas hidrosolubles, o solubles en agua, está compuesto por las vitaminas tipo B, o complejo vitamínico B, formado por las vitaminas B1 (tiamina), B2 (riboflavina), B3 (niacina o ácido nicotínico), B5 (ácido pantoténico), B6 (piridoxina), B7/B8 (biotina), B9 (ácido fólico), B12 (cobalamina) y vitamina C (ácido ascórbico). La adquisición de este tipo de vitaminas depende de la alimentación o dieta que tengamos, pudiendo incorporarse en nuestra ingesta por ejemplo a través de caldos y sopas resultantes de la cocción de ciertos alimentos, ya que se disuelven en el agua. Puesto que no se almacenan en el cuerpo (a excepción de la vitamina B12, que sí que lo hace en el hígado), es imprescindible su ingesta diaria y su exceso se elimina o excreta a través de la orina. En general, las vitaminas hidrosolubles son importantes en las diferentes reacciones y procesos del metabolismo del ser humano. Posteriormente en el capítulo (epígrafe 76.3) destacaremos de entre las vitaminas tipo B (importantes a nivel de hidratos de carbono, grasas y proteínas, así como en reacciones relacionadas con el sistema nervioso), la importancia y funciones de las vitaminas B9 y B12, así como la vitamina C.

Por su parte, el grupo de las vitaminas liposolubles, o solubles en lípidos, está compuesto por la vitamina A (retinol), vitamina D (colecalciferol), vitamina E (tocoferol) y la vitamina K (menadiona o antihemorrágica). Este tipo de vitaminas se adquieren a través de alimentos que contienen grasas y se almacenan en el cuerpo humano (hígado y tejidos grasos), en mayor o menor medida, sin ser excretadas o eliminadas a través de la orina. Esto implica que no es necesaria su ingesta diaria, pudiendo ser tóxica una ingesta 10 veces superior a la dosis o cantidad recomendada.

76.3 Las vitaminas que más nos suenan.

A continuación, nos adentraremos un poco más en una serie de vitaminas que son más conocidas dadas sus funciones e implicaciones en los procesos de nuestro cuerpo humano.

La vitamina B9 o ácido fólico (contenida en vísceras de animales, verduras de hoja verde, legumbres y frutos secos, entre otros), interviene en la formación de proteínas, leucocitos y eritrocitos, y es muy efectiva en el tratamiento de ciertas anemias. Además, suele ser prescrita a las mujeres antes de quedarse embarazadas ya que ayuda a prevenir deformaciones del feto, como pueden ser las relacionadas con la médula espinal o del cerebro.

A partir de fuentes animales como hígado, carne, pescado, huevos o leche adquirimos la vitamina B12 o cobalamina. Aunque solo necesitamos cantidades muy pequeñas de ella, esta vitamina resulta necesaria en el proceso de formación de proteínas y glóbulos rojos y en el correcto funcionamiento del sistema nervioso. En este caso, cabe señalar que como los vegetarianos no consumen animales o derivados de

éstos, suele ser aconsejado para ellos tomar suplementos en vitamina B12, para así evitar su déficit y por ende problemas de producción de glóbulos rojos, mielina o pérdida del epitelio, entre otros síntomas.

La vitamina C o ácido ascórbico (contenida en cítricos, fresas, pomelo, piña, brécol, tomates, espinacas, col, pimientos verdes o repollo, entre otros), por su parte, presenta un papel clave en la formación y conservación de colágeno en el cuerpo, en el metabolismo de proteínas y un importante papel antioxidante y cicatrizante. Su déficit prolongado en el cuerpo puede llegar a provocar la enfermedad del escorbuto. Sin embargo, debemos ser cuidadosos, ya que, aunque se elimina o excreta fácilmente por la orina, su exceso en el cuerpo puede provocar cálculos en vejiga y riñón, destrucción de vitamina B12 y pérdida de calcio en los huesos.

La vitamina D, o también llamada vitamina del Sol, está implicada en la formación de los huesos, además de en la adsorción de calcio y fósforo en el intestino. Aunque de forma mayoritaria esta vitamina se fabrica por el propio cuerpo cuando los esteroles, que se encuentran en una gran cantidad de alimentos, se desplazan a la piel y ésta recibe irradiación del Sol (esta es la razón por la que los médicos nos indican que tomar el Sol, de manera controlada, es bueno), también podemos adquirir una pequeña cantidad directamente a través de alimentos como la yema del huevo, hígado, atún o leche. En este caso, su deficiencia ocasiona raquitismo, osteoporosis o procesos inflamatorios; pero un consumo excesivo de ella puede provocar intoxicación en el cuerpo, junto con daños en el riñón, entre otras consecuencias.

La vitamina E, por su parte, podemos encontrarla en aceites vegetales, germen de trigo, hígado, yema de huevo y, en general, en estructuras con abundantes ácidos grasos. Entre las funciones que tiene, la de antioxidante biológico presenta un papel muy importante en la protección del desarrollo de enfermedades relacionadas con procesos oxidativos, como son las enfermedades cardiovasculares, diabetes, cáncer o los propios procesos de envejecimiento. Pero, además, interviene en la formación del DNA, RNA, glóbulos rojos, músculos y otros tejidos. Su deficiencia en el cuerpo humano puede producir parálisis o distrofia muscular y su exceso parece ser menos tóxico que otras vitaminas liposolubles.

La vitamina K es el elemento necesario y fundamental para que se produzca la coagulación sanguínea en nuestro cuerpo. La cantidad presente en nuestro cuerpo proviene de la sintetizada por las bacterias de nuestro intestino junto con la que obtenemos a través de todas las verduras de hoja verde, yema de huevo, hígado, aceite de soja o la propia soja. Si la cantidad de esta vitamina en nuestro cuerpo es muy escasa, podemos sufrir hemorragias dado que la sangre no se coagula.

Finalmente, no debemos olvidar la vitamina A, presente en alimentos como la zanahoria, melón, tomates o espinacas, entre otros, y que presenta un rol importante en los procesos de visión (de ahí que nos indiquen tomar zanahoria a los que no vemos adecuadamente).

76.4 Vitaminas e investigación.

Tras analizar las diferentes vitaminas, podemos decir que estos elementos orgánicos no son tan complejos como las proteínas o el DNA, pero tampoco tan

elementales como los oligoelementos y, sin embargo, son fundamentales o esenciales para el correcto funcionamiento del organismo. De hecho, cada día son más las enfermedades que se diagnostican fruto de las cantidades excesivas o defectuosas de vitaminas en el cuerpo; y es en esta dirección en la que se orientan las investigaciones actuales en este ámbito, intentando analizar el rol que las vitaminas presentan tanto en la promoción de la salud como en la prevención de enfermedades que se cree que están relacionadas con las vitaminas. Así por ejemplo *The American Journal of Clinical Nutritionlt* califica la deficiencia de vitamina D como un problema a escala mundial, y más común de lo que solemos pensar, lo que lleva a grupos de investigación, como los de la Universidad Estatal de Pennsylvania, a estudiar la relación entre la falta de vitamina D y determinadas enfermedades y procesos inflamatorios, así como con la exposición a la radiación solar.

Así mismo, existen grupos de investigación interesados en encontrar una relación entre los niveles de algunas vitaminas y enfermedades de deterioro cognitivo y/o la enfermedad de Alzheimer ya que existen indicios de existencia de niveles bajos o déficit de vitaminas en personas que sufren estas enfermedades, así como que algunos trastornos parecen mejorar aplicando vitaminas como tratamiento.

Hasta ahora, lo recomendado es tener una buena educación nutricional, seguir una dieta equilibrada y en casos concretos hacer uso de suplementos vitamínicos, por tiempos concretos, y siempre que éstos sean prescritos por nuestro médico. Sin embargo, este campo de investigación aún está por profundizar y explorar, siendo las próximas décadas muy importantes en la obtención de resultados y conclusiones.

Bibliografía:
¨Aspectos básicos y determinación de las vitaminas antioxidantes E y A¨. Mercedes Márquez, Carmen E. Yépez, Rosalía Sútil-Naranjo y Manuel Rincón (2002). E. Investigación Clínica, 43(3), 191-204.
¨La importancia de las vitaminas en la nutrición de personas que realizan actividad físico deportiva¨. Pardo Arquero, V. P. (2004). Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 4 (16) pp. 233-242.
¨Deficiencia de vitaminas y enfermedad de Alzehimer¨. Yeneisy Lanyau Domínguez y Consuelo Macías Matos (2005). Rev Cubana Salud pública, 31(4):319-26.
¨Vitamina D: evidencias y controversias¨ (2012). Infac. Liburukia, Vol. 20, Nº 2.

Mayte Conejero Muriel
Doctora Cristalización y Cristalografía
Investigadora Pos-doctoral UGR-CSIC

Capítulo 77
¿Cómo empezó la vida?
(Por Yaiza M. Castillo de la Peña)

¿Cómo empezó la vida en la Tierra? Para intentar responder a esta pregunta hay que remitirse a las tres principales teorías que intentan responderla: la sopa primigenia; las panspermias; y las fumarolas oceánicas. No obstante, de las tres, la primera es la que se encuentra actualmente más extendida.

77.1 Sopa primigenia

En 1924 el bioquímico Ivánovich Oparín postuló como se pudo haber originado la vida en una atmósfera reductora como se piensa que era la Tierra por entonces. En esta atmósfera predominarían las moléculas de carácter reductor (con menos oxígeno y más hidrógeno) como el metano (CH_4) y el amoníaco (NH_3), también moléculas de dióxido de carbono (CO_2), nitrógeno (N_2), agua (H_2O), sulfuro de hidrógeno (H_2S), y trazas de dihidrógeno (H_2). Gracias a la acción de las descargas eléctricas, la radiación solar, y la acción volcánica se podrían haber formado, a partir de ellas, sustancias orgánicas simples, que se fueron acumulando en los mares en formación (de aquí el nombre de sopa o caldo primigenios). Dado que la atmósfera no contenía moléculas de oxígeno (O_2), este no oxidaba a las moléculas orgánicas que se estaban formando, lo cual favoreció su formación.

Estas sustancias orgánicas se fueron acumulando en los mares, aumentando cada vez más y más su concentración, lo que provocó la formación de macromoléculas de mayor complejidad estructural, dando lugar a microgotas protegidas por polímeros, que albergarían en su interior las primeras enzimas. Estas macromoléculas cada vez se hicieron más complejas, aparecieron las primeras moléculas de ADN que codificaban para un genoma primitivo con capacidad para generar copias suyas exactas, y las primeras proteínas. Las cubiertas de las microgotas a su vez evolucionaron hacia membranas con capacidad de mantener su medio interno separado del externo, que además permitían la división. Se iniciaría así la primera célula.

En 1953, Stanley Miller y Harold Urey (bioquímico y químico estadounidenses, respectivamente), llevaron a cabo un experimento donde simularon las posibles condiciones de la atmósfera primitiva terrestre (hace 4000 millones de años). El objetivo del experimento era descubrir si se pueden generar moléculas complejas a partir de moléculas más sencillas. Dispuso en dos matraces de vidrio, en uno de ellos estaban los gases que conformarían la atmósfera primigenia, y en el otro calentaban agua para generar vapor con la finalidad de que estos gases recirculasen. En el matraz de los gases añadieron electrodos que generaban descargas eléctricas. Ambos matraces

estaban conectados y se disponía de un condensador donde recuperar muestras líquidas para analizar su composición.

Después de varias semanas de descargas, lograron encontrar aminoácidos, urea y ácidos grasos que se encuentran comúnmente en los seres vivos. Demostrando que sí que es posible que la vida en la Tierra se iniciase de esta forma.

77.2 Panspermias

En 1908 August Arrhenius propuso la teoría de las Panspermias (del griego *pan* "todo" y *sperma* "semilla"). En su hipótesis las Panspermias serían esporas (estructuras latentes y resistentes que adoptan algunas bacterias cuando no se dan las condiciones óptimas para su supervivencia y "reproducción") que se encuentran en el universo, y por ende en la superficie de meteoritos. Cuando estos meteoritos impactaron contra la Tierra, encontraron en ella las condiciones idóneas para su activación y "reproducción", iniciando así la vida en el planeta.

Esta teoría se sostiene debido a que gran cantidad de los aminoácidos que obtuvieron Miller y Urey en su experimento fueron encontrados en la superficie de un meteorito caído en Australia en 1969. No obstante, aunque esta teoría fuera cierta, solo nos está explicando el origen de la vida en la Tierra, pero no como se llegó a formar esta vida.

77.3 Fumarolas oceánicas

La tercera teoría apareció a raíz de hallazgos recientes gracias a la oceanografía. En 1977 una campaña oceanográfica norteamericana descubrió, en el archipiélago de las Galápagos, a más de 2500m de profundidad todo un nuevo ecosistema, con organismos desconocidos hasta la fecha. Lo sorprendente de este descubrimiento es que, todos los ecosistemas, tanto acuáticos como terrestres, dependen de organismos autótrofos. Estos organismos son capaces de generar su propia materia orgánica a partir de dióxido de carbono y luz. ¿Qué hay de extraño en este caso? La luz a 2500m es nula. Años después se descubrió que existen bacterias que no utilizan la luz del sol para sintetizar la materia orgánica, sino la oxidación del sulfuro de hidrógeno, y resulta que las fumarolas oceánicas son unas altas productoras de este compuesto.

Las fumarolas oceánicas se originan cuando el magma volcánico asciende y luego solidifica. Al solidificarse este se resquebraja, el agua de mar se infiltra calentándose a altas temperaturas (300-400ºC) y sale por las fisuras. Esta agua marina en evaporación es carente de oxígeno, está muy acidificada, presenta altas cantidades de minerales y lleva compuestos tóxicos como el sulfuro de hidrógeno. El fluido hidrotermal en contacto con el agua de mar, forma estas fumarolas que llegan a alcanzar los 20 metros de altura, creando vida a su alrededor. ¿Pudo haber sido este el inicio de la vida en la Tierra?

Aún sin conocer a ciencia cierta cuál es el origen de la vida en la Tierra, y sin tener una respuesta exacta, sí que tenemos registros fósiles de unas de las estructuras más antiguas de vida terrestre. Estos registros se encuentran en rocas halladas en Australia y Sudáfrica, datan de unos 3500 millones de años y presentan cierta similitud

con las cianobacterias filamentosas actuales. No sabemos si fueron los primeros, ni de donde proceden, pero lo que nos indican es que la vida en la Tierra se originó unos 600 millones de años después de la creación del planeta.

Como no, esta es una investigación que sigue abierta y que nos costará muchos años desentrañar, si es que alguna vez logramos descubrirlo. Para ello hay que seguir investigando y contrastando resultados. Es un área de estudio que abarca muchas disciplinas: desde la física, la astronomía, la oceanografía, pasando por la microbiología, química, bioquímica, etc. pero lo que sí es seguro, por ese mismo motivo, es que entre todos podemos aportar conocimientos que nos ayuden a descubrir: como apareció la vida en el planeta Tierra.

Bibliografía:
"Origen de lo viviente" (2015). Sección i 1. Ediciones Santillana

Yaiza M. Castillo de la Peña
Estudiante de doctorado
Institut de Ciències del Mar - CSIC

Capítulo 78
¿Cómo se descubrieron los virus?

(Por Yaiza M. Castillo de la Peña)

Para poder entender cómo se llegaron a descubrir los virus, primero tenemos que saber que son. Los virus (del latín *virus*, «toxina» o «veneno») son unos organismos microscópicos que solo pueden replicarse dentro de un hospedador. Esto se debe a que los virus están formados solo por un ácido nucleico recubierto por una envuelta proteica (llamada cápside), que puede, o no, estar envuelta a su vez de una cubierta lipídica. Esto hace que los virus no presenten un metabolismo propio, y requieran de la maquinaria de replicación de una célula hospedadora que trabaje para ellos.

La mayoría de los virus son tan pequeños que solo pueden ser observados con la ayuda de un microscopio (de epifluorescencia, o para mayor amplificación uno electrónico). No obstante, recientemente se están descubriendo virus “gigantes”, que llegan a superar en tamaño a algunas bacterias y pueden observarse usando un microscopio óptico convencional.

Hay millones de tipos diferentes de virus, e infectan a todos los tipos de organismos: animales, plantas, bacterias y archaeas. Se encuentran prácticamente en todos los ecosistemas de la Tierra y se cree que son las entidades biológicas más abundantes del planeta.

¿Cómo se llegaron a descubrir? ¿Cuál fue el primero?

Para ello hay que tener en cuenta dos momentos en la historia. El primero: cuando se detectó por primera vez una partícula infecciosa nueva de tipo no bacteriano, y el segundo: cuando se describió y nombró por primera vez a esa partícula infecciosa como *virus*.

La primera vez que se tuvo consciencia de que una partícula infectiva, no bacteriana, estaba produciendo una infección, fue en 1892. Dmitri Ivanoski, biólogo ruso, estaba estudiando extractos de hojas de tabaco infectadas cuando se percató de que después de la filtración del extracto, este seguía siendo infeccioso. La intención de Ivanoski, podemos suponer, era aislar la “bacteria” que estaba afectando a las hojas de la planta del tabaco, ya que hasta entonces el único agente microbiano infeccioso del que se tenía conocimiento eran las bacterias. Para ello usó un filtro de Chamberland, donde las bacterias se retienen y se consigue un extracto limpio de infección. No obstante, en el filtrado del extracto de Ivanoski aparecía algo que seguía infectando a las hojas de tabaco. ¿Qué tuvo de importante este resultado? Se estaba descubriendo un patógeno filtrable.

Seguramente Ivanoski no llegó a comprender completamente su descubrimiento, pero sí lo hizo Martinus Willem Beijerinck en 1898. Beijerinck, naturalista, botánico, y microbiólogo neerlandés, fue el primero en postular que un "contagium vivum fluidum" y, por tanto, no un "fixum contagium" como son las bacterias, era el agente causante de la enfermedad del mosaico del tabaco. ¿Cómo llegó Beijerinck a esta conclusión? A partir de un experimento con geles de agar. Beijerinck observó que esta nueva partícula infecciosa era capaz de migrar en un gel de agar, tratándose por tanto de un agente infeccioso soluble. Asimismo, Beijerinck también observó que este agente soluble era capaz de multiplicarse cuando estaba en contacto con el metabolismo del huésped, y de que era capaz de distribuirse en los vasos del floema junto con los nutrientes de la planta, hechos que no se correspondían con la teoría microbiana prevalente en ese momento.

Por tanto, el virus del mosaico del tabaco jugó un papel clave en el origen de la virología tal y como la conocemos ahora. Desde Ivanosky hasta Beijerinck, pasando por el también científico Adolf Mayer, el concepto de agente filtrable, demasiado pequeño para ser observado en un microscopio óptico, pero capaz de causar una enfermedad al multiplicarse en las células vivas y en los tejidos, había nacido. Se habían descubierto los virus.

En ese mismo año (1898) Loeffler y Frosch rápidamente describieron y aislaron el primer virus animal, causante de la fiebre aftosa del ganado, y en 1901 Walter Reed y su grupo reconocieron el primer virus humano, el virus de la fiebre amarilla. Con lo cual, a principios del siglo XX, el concepto de virus ya estaba firmemente establecido.

Hoy en día, el estudio de los virus está cada vez más en auge. Nos hemos dado cuenta de que los virus juegan roles muy importantes en todo lo que nos rodea: desde causar enfermedades hasta controlar los flujos de carbono en el océano. No solamente existen virus que infectan a animales (entre ellos los humanos), o plantas, también los hay que infectan a bacterias y archaeas (llamados fagos), lo que nos puede llevar al debate sobre el uso de los virus como terapia fagica. Asimismo, recientemente se están descubriendo los virus "gigantes" de los que hablábamos al principio, así como virus que se han mantenido congelados e inactivos durante miles de años en el permafrost y que están actualmente siendo estudiados.

La virología es una gran área de estudio de la que queda mucho por conocer y de la que esperamos seguir aprendiendo muchas cosas.

Bibliografía:
Bos, L. (1999). Beijerinck's work on tobacco mosaic virus: historical context and legacy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 354(1383), 675–685.
Lecoq, H. (2001). Discovery of the first virus, the tobacco mosaic virus: 1892 or 1898? *C R Acad Sci III*. 324(10):929-33.
Lustig, A., & Levine, A. J. (1992). One hundred years of virology. *Journal of Virology*, 66(8), 4629–4631.

Yaiza M. Castillo de la Peña
Estudiante de doctorado
Institut de Ciències del Mar - CSIC

Capítulo 79
¿Es posible una vida basada en el silicio?
(Por David L. Van Vranken y Vanessa Arredondo)

Los átomos de carbono y moléculas que contienen enlaces carbono-carbono son fundamentales para todos los organismos vivos de la Tierra. Por ejemplo, constituyen la estructura del ADN, ARN, de proteínas, plantas, animales, seres humanos, y mucho más. Aun así, ¿será posible que la vida se centre alrededor de un elemento además del carbono? La tabla periódica, introducida en la década de 1860, ordena los elementos por tamaño y propiedades químicas. Los elementos en la misma columna anticipan a exhibir propiedades químicas similares. Silicio debería ser el elemento químicamente más parecido al carbono ya que se encuentra directamente debajo del carbono en la tabla periódica. Ambos tienen cuatro electrones de valencia en su capa exterior que permite a los dos unirse con cuatro grupos para formar estructuras tetraédricas.

Sin embargo, hay algunas diferencias importantes entre los enlaces que el carbono y el silicio pueden hacer. Enlaces de carbono-carbono son un poco más fuertes que los enlaces de silicio-silicio debido al tamaño más grande del átomo de silicio. El carbono está limitado por la regla del octeto – solo puede tener ocho electrones de valencia – y por lo general se limita a cuatro enlaces. Silicio no está limitado por la regla del octeto y puede tener más de ocho electrones de valencia. Silicio prefiere cuatro enlaces, pero fácilmente produce moléculas estables en las que tiene cinco o incluso seis enlaces con otros átomos. La habilidad para formar más enlaces debería permitir al silicio producir una mayor variedad de estructuras moleculares complejas comparado con el carbono. Además, el silicio predomina en cantidad sobre el carbono en planetas rocosos como la Tierra. El silicio constituye el 14% de la Tierra por peso mientras el carbono representa menos del 0.1%. ¿Dada la promiscuidad química y la abundancia de silicio podemos visualizar formas de vida basadas en el silicio en lugar del carbono?

La cuestión sobre vida basada en el silicio no es nueva. Muchos han especulado sobre las formas de vida basadas en el silicio. La educación bioquímica de Isaac Asimov y su visión diletante lo llevó a evaluar las características químicas del carbono en relación con el silicio. Ambos elementos forman óxidos estables con oxígeno molecular: el carbono forma dióxido de carbono (CO_2) y el silicio forma sílice (Si_nO_{2n}). Las propiedades físicas de estos dos compuestos difieren dramáticamente. El dióxido de carbono se procesa sin esfuerzo; es un gas que se disuelve fácilmente en agua y reacciona con el agua. En contraste, sílice rápidamente forma sales de silicato refractario. Sílice y silicatos son el principal componente de la roca; son insolubles en agua y no son reactivos con agua. Además, la mayor estabilidad de sílice en relación con la del dióxido de carbono hace que sea más difícil de reducirlo para formar enlaces útiles como silicio-hidrógeno o silicio-silicio. La vida en nuestro planeta últimamente es posible por la reducción bioquímica de CO_2 para formar moléculas compuestas de enlaces carbono-hidrógeno y carbono-carbono.

La vida también depende de una serie de otros procesos bioquímicos tales como la replicación, la adaptación, y el metabolismo. Estos procesos son dependientes de las interacciones que involucran anillos planos que son un átomo grueso. La forma plana de estos fragmentos moleculares depende de la capacidad de formar enlaces dobles estables y planos. Los átomos de carbono, nitrógeno, y oxigeno pueden hacerlo. Además, estos elementos también pueden combinarse para formar enlaces dobles estables de carbono-nitrógeno, carbono-oxígeno, y nitrógeno-oxígeno (Figura 1). Los átomos de silicio no pueden formar enlaces dobles estables entre sí o con otros átomos. Enlaces dobles de silicio-silicio son más débiles que enlaces dobles típicos de carbono-carbono (14-24 kcal/mol comparado a más de 48 kcal/mol) y requieren extremadamente grandes grupos en el átomo de silicio para prevenir reacciones espontáneas con el oxígeno o el agua. Además, enlaces dobles de silicio-silicio no son planos debido a que los átomos de silicio prefieren geometrías piramidales (Figura 1).

Figura 1: *Ejemplos de enlaces dobles estables que incluyen carbono y de un enlace doble inestable del silicio.*

La capacidad del carbono para formar moléculas planas estables con otros átomos de carbono hace que sea un átomo potente. En una de sus formas más puras, grafito, átomos de carbono forman enlaces con otros tres átomos de carbono. En esta disposición, los átomos de carbono forman láminas u hojas planas (Figura 2, izquierda). En una representación simplificada, cada átomo de carbono está enlazado doblemente a otro átomo de carbono. Estos dobles enlaces son conocidos como enlaces π (pi). Las láminas bidimensionales de átomos de carbono en el grafito se apilan unas encima de las otras. Cuando se apilan, los enlaces pi de una hoja interactúan con los enlaces pi de la otra hoja. Esas interacciones se conocen como apilamiento pi, y son estables; hacen que el agregado de hojas sea más fuerte que las hojas individuales. El diamante es otra forma pura de carbono, cada átomo de carbono está unido a otros cuatro átomos de carbono y forman una red tridimensional rígida (Figura 2, derecho). En condiciones normales, el grafito es ligeramente más estable que el diamante, debido a las favorables interacciones de apilamiento pi entre las hojas. La energía de estas interacciones entre las láminas de grafito es de aproximadamente 1.4 kcal/mol. Esto significa que las hojas apiladas de carbono son diez veces más estables que las hojas aisladas.

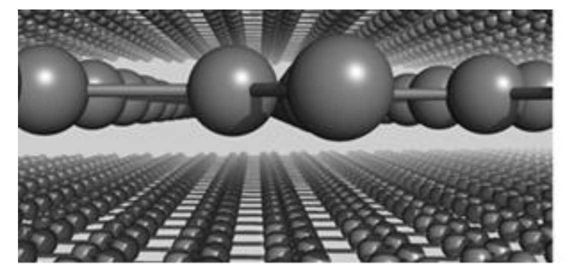
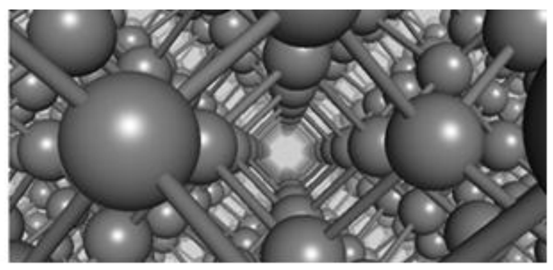



Figura 2: *Izquierda: En grafito, los átomos de carbono están ordenados en láminas apiladas. Derecha: En diamante, los átomos de carbono están ordenados en una red tridimensional.*

Apilamiento pi es igualmente posible cuando las moléculas planas contienen enlaces dobles entre carbono, nitrógeno, y oxígeno en diversas combinaciones. Las moléculas de la vida son completamente dependientes de las interacciones del apilamiento pi entre componentes atómicamente planos. Por ejemplo, en el ADN, el apilamiento pi de subunidades llamadas bases es esencial para el auto-ensamblaje de las unidades en la estructura icónica doble hélice (Figura 3, izquierda). Las bases de ADN son anillos planos compuestos de átomos de carbono, nitrógeno y oxígeno. Las interacciones de apilamiento pi entre las bases de ADN son esenciales para la estabilidad del ADN y para su función como repositorio de información genética. La molécula biológica ARN también contiene bases y es ampliamente reconocido como el precursor molecular a ADN en la génesis de la vida. No es sorprendente que interacciones similares de apilamiento pi son evidentes en la estructura química del ARN (Figura 3, centro). Por último, la inspección de las proteínas revela la importancia de las interacciones de apilamiento pi, tanto dentro de la estructura plegada de proteínas como entre las proteínas y su par de reacción (Figura 3, derecha).

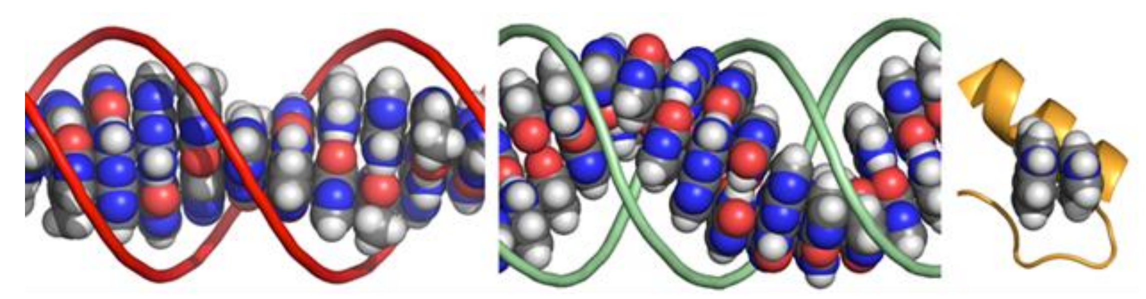


Figura 3: *Interacciones apilamiento pi entre las bases planas en las moléculas de la vida. Izquierda: bases planas apiladas en el ADN; Centro: bases planas apiladas en ARN; Derecha: cadenas laterales de proteínas se apilan en una enzima metabólica.*

El carbono puede formar enlaces dobles planos y estables y también anillos planos que son necesarios para los procesos de reconocimiento molecular responsable de la vida. El silicio no forma enlaces dobles estables con cualquier átomo: nitrógeno, carbono, oxígeno o incluso a sí mismo. Dado que el silicio no puede formar enlaces dobles estables, no puede formar estructuras atómicamente planas capaces de apilamiento. Un universo bioquímico basado en silicio carecería de la capacidad para el reconocimiento molecular cara a cara a través de apilamiento pi. Sería semejante a un almacén que contiene envases irregulares que no podían ser apilados. Un sistema bioquímico basado en silicio se enfrenta a un reto de diseño serio que no es fácil de superar. Aunque todas las condiciones ambientales necesarias existirían para permitir enlaces con silicio como el constituyente principal de los procesos bioquímicos, silicio siempre se quedará corto. En un mundo donde la vida depende de los procesos moleculares que implican estructuras atómicamente planas, el silicio nunca será capaz de apilar.

Bibliografía:
1. a. Asimov, I. "Big Brother" *The Magazine of Fantasy and Science Fiction*. September, 1982. b. Asimov, I. "V. Big Brother" *X Stands for Unknown*. Doubleday: New York, 1984. p. 61-71.
2. a. Asimov, I. "Bread and Stone" *The Magazine of Fantasy and Science Fiction*. October, 1982. b. Asimov, I. "VI. Bread and Stone" *X Stands for Unknown*. Doubleday: New York, 1984. p. 72-82.
3. a. Asimov, I. "VII. A Difference of an E" *The Magazine of Fantasy and Science Fiction*. November, 1982. b. Asimov, I. "VII. Big Brother" *X Stands for Unknown*. Doubleday: New York, 1984. p. 83-94.
4. a. Asimov, I. "Silicon Life After All" *The Magazine of Fantasy and Science Fiction*. December, 1982. b. Asimov, I. "VIII. Silicon Life After All" *X Stands for Unknown*. Doubleday: New York, 1984. p. 95-108.
5. Iwamoto, T.; Ishida, S. "Multiple Bonds with Silicon: Recent Advances in Synthesis, Structure, and Functions of Stable Disilenes" in *Functional Molecular Silicon Compounds II*. Scheschkewitz, D., Ed. Structure and Bonding, Vol 156. Springer: Switzerland, 2013. pp. 125-202.
6. Zacharia, Renju "Chapter 4. Energetics of interlayer binding in graphite." in *Desorption of Gases from Graphitic and Porous Carbon Surfaces*. Dissertation. Freie Universtitat Berlin. 2004.
7. Girifalco, L. A.; Lad, R. A. Lad, *J. Chem. Phys.*, 1956, 25, 693.
8. Winter, N. W.; Ree, F. H. "Stability of the Graphite and Diamond Phases of Finite Carbon Clusters" *Detonation Symposium* Snowmass, CO August 30 – September 4,1998.

David L. Van Vranken — Doctor en Química, Profesor Titular
Vanessa Arredondo — Candidata a Doctora en Química
Universidad California de Irvine

Capítulo 80
¿Por qué se extinguieron los Dinosaurios?
(Por Pedro Pereda Gómez)

Los dinosaurios aparecieron hace unos 230 millones de años, durante el periodo Triásico, y dominaron la tierra, el aire y los mares de nuestro planeta durante más de 150 millones de años. Durante este periodo evolucionaron algunas de las criaturas más impresionantes de la naturaleza: dinosaurios carnívoros de 6 metros de altura, dinosaurios herbívoros de 26 metros de longitud y pterosaurios con la envergadura de 12 metros.

Sin embargo, hace 65 millones de años se extinguió el último dinosaurio no aviario. Igual que los gigantescos mosasaurios y plesiosaurios que en su día dominaron los mares, y los pterosaurios en los cielos. Pero no solo los dinosaurios, también el plancton, la base de la cadena alimenticia del océano, se vio muy afectado. Además, muchas familias de braquiópodos y esponjas de mar desaparecieron, igual que los restantes ammonites de concha dura. También se redujo la gran diversidad de tiburones y se marchitó la mayor parte de la vegetación. En resumen, se eliminó más de la mitad de las especies mundiales.

¿Qué causó esta masiva extinción que marca el final del Cretácico y el comienzo del Paleógeno?

Los científicos han coincidido en dos hipótesis para explicar la extinción del Cretácico: un impacto extraterrestre, por ejemplo un asteroide o un cometa, o un período de gran actividad volcánica. Cualquiera de los dos escenarios habría ahogado los cielos con restos que privaron a la Tierra de la energía del sol, impidiendo la fotosíntesis y extendiendo la destrucción arriba y abajo de la cadena alimenticia. Una vez que se asentó el polvo, los gases de efecto invernadero bloqueados en la atmósfera habrían provocado que se disparara la temperatura, un repentino cambio climático acabó con mucha de la vida que logró sobrevivir a la prolongada oscuridad.

La teoría del impacto extraterrestre se basa en el descubrimiento de un estrato o capa de arcilla de color oscuro en los sedimentos que datan de la época de la extinción. Este estrato tiene, entre otras características, un alto nivel de Iridio, elemento químico raro en la corteza terrestre que, sin embargo, se encuentra en los meteoritos con la misma concentración que en este estrato. Puesto que este estrato se encuentra en todo el planeta, en la tierra y en los océanos, esto condujo a los científicos a afirmar que el Iridio se esparció por el planeta cuando un cometa, o un asteroide, cayó en algún lugar de la Tierra y a continuación se evaporó. Este estrato también presentaba abundancia de hollín, el cual debió producirse por el gran incendio que siguió a la caída del meteorito.

Los defensores de la teoría de la gran actividad volcánica nos indican que el núcleo de la Tierra también es rico en Iridio, y el núcleo es el origen del magma que vomitó la Tierra en la Meseta del Decán (India), una de las mayores formaciones volcánicas de la Tierra, donde se apiló material volcánico sobre más de 1.500.000 kilómetros cuadrados de superficie y 2000 metros de espesor. Este período de actividad volcánica también se ha calculado que ocurrió entre 60 y 68 millones de años, al final del Cretácico, y habría extendido el Iridio por todo el planeta, junto con el polvo que ocultaba la luz solar y los gases de efecto invernadero.

Sin embargo, la teoría del impacto extraterrestre se vio reforzada por otra característica de este estrato, que es la gran acumulación de esférulas de vidrio, que se producen cuando se solidifica rápidamente roca vaporizada, y de un tipo especial de cristales de cuarzo llamado "cuarzo de impacto" que solo se encuentra en las proximidades de los cráteres producidos por la caída de objetos del espacio.

Pero si era cierto que hace 65 millones de años un meteorito de gran tamaño, capaz de ser el responsable de la extinción masiva descrita, había caído en la Tierra, era necesario localizar el lugar del impacto.

Puesto que no se conocía en la superficie continental ningún cráter que se adaptara a las características de este meteorito, se partió de la posibilidad de que el impacto hubiera ocurrido en el océano. Partiendo de esta hipótesis, los científicos buscaron las huellas que hubiera dejado un tsunami posterior al impacto. Estas huellas, denominadas tsunamitas, son unos depósitos caracterizados por una sedimentación caótica y se encontraron en Texas (Estados Unidos), México y en diversas localidades del Caribe. Finalmente, en el año 1991 se descubrió el cráter de impacto en Chicxulub, en la península de Yucatán (México).

El cráter de Chicxulub, cuya antigüedad se ha fijado en 65 millones de años, mide más de 180 kilómetros de diámetro y forma una de las zonas de impacto más grandes del mundo. Se estima que el "bólido" que lo formó medía al menos 10 kilómetros de diámetro. Este cráter se encuentra bajo más de un kilómetro de sedimentos carbonatados, lo que ha desfigurado su aspecto topográfico y ha hecho difícil su localización, pero los estudios en profundidad realizados permitieron descubrir el contorno del cráter.

Se estima que el impacto del "bólido" fue dos millones de veces más potente que el dispositivo explosivo más potente creado por el hombre, jamás detonado, con una potencia de 50 megatones. Incluso la mayor erupción volcánica explosiva que se conoce en la historia reciente, la que creó la Caldera de la Garita en Colorado (Estados Unidos) hace entre 40 y 25 millones de años, fue significativamente menos potente que el impacto de Chicxulub.

El impacto habría causado algunos de los megatsunamis más grandes de la historia de la Tierra. Una nube de polvo, cenizas y vapor habrían extendido el diámetro y área del cráter, cuando el meteorito se hundía en la corteza terrestre en menos de un segundo. El material excavado, junto con trozos del asteroide, habrían sido eyectados a la atmósfera por la explosión, se habrían calentado hasta convertirse en cuerpos incandescentes y habrían reentrado a la propia atmósfera terrestre, quemándola y posiblemente provocando incendios globales. Mientras tanto, enormes ondas de choque habrían causado terremotos y erupciones volcánicas globales. La emisión de

polvo y partículas podrían haber cubierto la superficie entera de la Tierra durante varios años, posiblemente una década, creando un medio de vida difícil para los seres vivos. La producción de dióxido de carbono provocada por el choque y por la destrucción de rocas carbonatadas habría causado un dramático efecto invernadero. Otra consecuencia del impacto es que las partículas de polvo de la atmósfera habrían impedido que la luz solar llegara a la superficie de la Tierra, disminuyendo la temperatura drásticamente. La fotosíntesis de las plantas habría quedado interrumpida, afectando a la totalidad de la red trófica.

El descubrimiento de otros cráteres, como el cráter de Shiva en el Océano Indico, el cráter Silverpit en el Mar del Norte frente a las costas del Reino Unido y el cráter Boltysh en Ucrania, probablemente causados por grandes objetos extraterrestres que impactaron contra la Tierra, condujo a la teoría del impacto múltiple por la cual Chicxulub solo fue uno de varios impactos que habrían ocurrido aproximadamente al mismo tiempo. La colisión del cometa Shoemaker-Levy 9 con Júpiter en 1994, en la cual el asteroide se fragmentó en varias partes antes de la colisión, demostró que las interacciones gravitacionales pueden afectar a un cometa creando la posibilidad de múltiples impactos en un período de días. Algo similar pudo ocurrir en la Tierra hace 65 millones de años.

En cualquier caso, ya sea por el impacto de uno o de varios meteoritos, la teoría del impacto extraterrestre goza actualmente de una aceptación amplia, aunque algunos científicos todavía la critican por considerar que de ser cierto un impacto de estas características, ¿por qué sobrevivieron mamíferos, tortugas, cocodrilos, salamandras y ranas? Las aves también se libraron, al igual que las serpientes, bivalvos y los erizos y estrellas de mar. Incluso a las plantas resistentes capaces de soportar climas extremos les fue bien.

Ambas teorías, la del impacto extraterrestre y la del vulcanismo masivo, son meritorias. Sin embargo, en Marzo de 2010, expertos de Europa, Estados Unidos, México, Canadá y Japón confirmaron que la extinción masiva fue originada por el impacto de un meteorito, quedando desvirtuadas otras hipótesis como la del vulcanismo masivo.

Independientemente de lo que provocó la extinción, ésta marcó el fin del reino de terror de los grandes carnívoros bípedos del Triásico, como el Tyrannosaurio rex, y permitió que los mamíferos se diversificaran rápidamente y evolucionaran a nichos recién abiertos.

Pero, ¿realmente han desaparecido todos los dinosaurios? La respuesta es no. Al principio hemos hablado de la extinción de los "dinosaurios no aviarios", porque los científicos consideran que las aves actuales son los supervivientes de un grupo de dinosaurios terópodos que surgieron durante la era Mesozoica y que lograron superar la catástrofe y evolucionar. Recordemos que también eran dinosaurios terópodos los grandes carnívoros bípedos del Triásico, aunque estos tuvieron peor fortuna.

Las aves comparten muchas características únicas, esqueléticas y de comportamiento, con los dinosaurios. Por otra parte, muchos fósiles de dinosaurios se han recogido con plumas preservadas, e incluso algunos tienen largas plumas en brazos y patas formando alas.

Por tanto podemos concluir que las aves de hoy en día son los seres más parecidos a los dinosaurios que habitaron la Tierra en el pasado. Solo necesitamos levantar la vista y mirar al cielo para ver los dinosaurios de hoy.

Pedro Pereda Gómez
Licenciado en Biología
Director de Hemasoft

Capítulo 81
¿Por qué persisten los recuerdos?
(Por José Viosca)

El primero fue en una playa. Estábamos de pie, en la arena, cuando un contacto de labios lo detuvo todo. Una ola, a punto de romper en la orilla, quedó congelada detrás de nosotros. Paró también el viento, dejando su larga melena ondeando y quieta como aquella bandera ondulada en la luna. En la boca, una lengua cálida y un sabor nuevo, intenso, abrasador. Escuché una melodía de violines y sentí unas mariposas revoloteando en el estómago. Entonces abrí los ojos y respiré. Vi frente a mí una sonrisa, y no supe si habían pasado diez segundos o diez horas.

Así fue el primer beso. Uno de esos momentos que uno recuerda con nitidez, como el nacimiento de los hijos y otros instantes únicos. Evocamos con facilidad dónde nos encontrábamos, con quién, qué ropa llevábamos. Si hacía calor o frío, el olor del lugar, y, por supuesto, la música ambiental (o los violines sonando en mi cabeza). El primer beso, en general cualquier recuerdo duradero, queda grabado en la memoria como si fuera un DVD que podamos luego reproducir a voluntad para recuperarlo (¿intacto?).

El paralelismo entre memoria y música nos puede ayudar mucho a entender cómo se las arregla el cerebro para almacenar los recuerdos. Ambos, memoria y música, tienen una base física. Si bien la música surge por las vibraciones de las moléculas del aire, los recuerdos, por su parte, se deben a cascadas de interacciones moleculares y flujos de átomos que tienen lugar en el interior del cerebro.

De hecho, el cerebro es una especie de orquesta. Una orquesta que utiliza un lenguaje propio y cuyos instrumentos son las neuronas, unas células que, cuando recordamos algo, no solo disparan ráfagas de impulsos nerviosos, sino que lo hacen de forma coordinada, miles de neuronas al mismo tiempo, como virtuosos músicos que ejecutan con precisión una partitura.

De esta forma, cada recuerdo es una especie de melodía. Una melodía única que, como toda sinfonía, tiene su propia partitura. Para cada memoria, el cerebro almacena un registro, a modo de archivo, para su consulta posterior. A cada memoria le corresponde un pentagrama que se guarda de forma dispersa en muchos lugares del cerebro, dentro de miles de neuronas que almacenan un trozo cada una. Cada neurona guarda y representa un matiz único, un aspecto diferente del recuerdo, así como cada instrumento de la orquesta lee una partitura distinta de la sinfonía. Y dentro de las neuronas, un conjunto de moléculas afina sus ráfagas de actividad, como se afinan los instrumentos de la orquesta. Una afinación que no solo sintoniza las neuronas sino que es la base material misma de la memoria.

El cerebro, el órgano donde residen los recuerdos, está hecho de células que a su vez están hechas de moléculas. Por eso mismo, la duración de los recuerdos plantea un auténtico rompecabezas biológico [1]. Por seguridad, para evitar que acumulen errores peligrosos como un coche viejo, casi todas las moléculas son destruidas pocas horas o días después de fabricarse. Sin embargo, los recuerdos duran años o, el del primer beso, toda la vida. ¿Cómo son capaces entonces las moléculas y neuronas del cerebro de almacenar los recuerdos?

81.1 La partitura (la traza de memoria)

Ya en su mitología, los griegos describieron la memoria como un proceso de inscripción. Mnemea, la musa de la memoria (también de la plasmación o la creación), aparece en obras de arte como una muchacha escribiendo. La joven apoya un estilete en su barbilla, intentando recordar lo que piensa para escribirlo en un rollo de papiro.

Inspirándose en esta idea, un zoólogo alemán llamado Richard Wolfgang Semon (1859-1918) conjeturó ya en el siglo XIX que la base cerebral de la memoria debía ser alguna forma de “traza” [2]. Se refería a un “registro permanente, escrito o grabado sobre la sustancia irritable” (el cerebro). 150 años después, sabemos que la traza de memoria está hecha en realidad de una variedad de componentes químicos y físicos.

En el cerebro, como en toda orquesta, hay un director que compone las partituras. Y también un vasto escenario con miles de instrumentos, cada uno con una función determinada. Hay regiones cerebrales dedicadas a codificar en el lenguaje neuronal cada parámetro del mundo sensorial, así como regiones dedicadas a enviar órdenes a los músculos para movernos. Hay neuronas que responden y representan conceptos, lugares en el espacio y formas geométricas; y neuronas cuya activación causa movimientos, miedo, satisfacción, bienestar, y otros procesos neurológicos.

El cerebro es una orquesta muy especial porque, constantemente, escribe y reproduce su música. Una música con la que representa internamente, como una especie de código cerebral, aquello que vivimos. (En realidad, aquello que interpretamos que vivimos). En cierto sentido, el compositor que hay en el cerebro anota en un diario nuestras vivencias (o aquello que interpretamos de nuestras vivencias), pero en vez de textos y fotos, el compositor escribe partituras. La memoria es ese diario que contiene todas las partituras de nuestra vida.

Solo que en el cerebro, no hay papel sino neuronas. Neuronas que son a la vez los instrumentos que ejecutan las sinfonías y el sustrato material del diario donde se escribe la partitura de la memoria.

81.2 El compositor (el hipocampo)

Sabemos por Henry Molaison (1926-2008), seguramente el paciente más conocido en la historia de la investigación de la memoria, que cada componente de la orquesta está situado en una región particular del cerebro. A los 27 años, a partir de una operación quirúrgica en la que quedaron lesionadas varias regiones de su cerebro, especialmente una zona llamada hipocampo, Molaison perdió la capacidad de formar memorias duraderas. Resultó extraordinario que el problema afectara solo a un tipo

concreto de memoria: las memorias episódicas (los recuerdos de momentos, lugares, personas). Durante décadas, Molaison vivió atrapado en el pasado y un presente efímero. Todo se desvanecía en su mente en pocos segundos: desconocía qué estaba haciendo, con quién hablaba, o si había almorzado [3]. Pero los recuerdos anteriores a la operación, también el de su primer beso, permanecieron intactos en otras regiones del cerebro. Gracias a Molaison, sabemos que el hipocampo es el compositor que escribe las partituras, pero el papel donde quedan grabadas las memorias son neuronas en otros lugares del cerebro. La biblioteca donde quedan archivados los recuerdos es la capa más externa del cerebro, la corteza cerebral [4].

Dentro de las neuronas, las trazas de memoria se guardan principalmente en las sinapsis (unas regiones con forma de botón abombado por donde las distintas neuronas se comunican entre sí). Así lo pensó por primera vez hace más de 100 años Santiago Ramón y Cajal [5], el primer español en recibir un Premio Nobel en una disciplina científica. Cajal acertó en su intuición, aun sin poder demostrar empíricamente que las sinapsis existían – eso ocurrió décadas más tarde. Hoy sabemos que cada neurona en la corteza cerebral humana conecta 10.000 veces con otras neuronas, configurando una enorme cantidad de redes neuronales susceptibles de almacenar recuerdos. Probablemente, el cerebro es la orquesta más grande del mundo.

Como si afinaran las distintas cuerdas de una guitarra o los distintos engranajes de cada instrumento de una orquesta, las neuronas almacenan memoria ajustando sus conexiones, ya sea potenciándolas o debilitándolas. De esa forma, pensó el psicólogo canadiense Donald Hebb (1904-1985) a mediados del siglo XX, las neuronas podrían establecer lazos de realimentación y mantener una actividad reverberante. Esa sería la base cerebral de la memoria, según Hebb, una idea que también tardarían años los científicos en demostrar empíricamente.

Hebb planteó su idea con un ejemplo concreto de aprendizaje: el condicionamiento de Pávlov (el de los perros que salivan al oír una campana), y un modelo minimalista con dos neuronas [6]. Hebb imaginó lo que debía suceder en el interior del cerebro de Tungus, una especie de pastor alemán y uno de los 50 perros que Pávlov estudió, cuando aprendía el condicionamiento del reflejo de la salivación. Al principio, la neurona S, que representa y desencadena la salivación y por ello dispara (uno o varios impulsos nerviosos) cuando Tungus ve comida, está desconectada de la neurona C, que representa el sonido y dispara cuando el perro oye una campana. Por tanto, la neurona S y la neurona C no pueden inicialmente sincronizarse. Pero si la comida y la campana son mostradas a la vez varias veces, sucede algo crucial: como resultado del disparo simultáneo, ambas neuronas refuerzan sus conexiones. Así, cuando más tarde la neurona C sea activada (porque suene la campana), C activará a S y Tungus salivará. En el cerebro de Tungus, se habrá escrito entonces una partitura en su diario de recuerdos.

81.3 La tinta (moléculas que perduran)

Salvador Dalí pintó su cuadro más famoso, la persistencia de la memoria, inspirándose en su tierra natal. Sobre un paisaje del Alto Ampurdán (Gerona), en el centro del lienzo pintó unos relojes deformados – él los llamó relojes blandos-, para expresar una voluntad de trascendencia, un deseo de perdurar en el tiempo más allá de su propia muerte. Dalí pintó el cuadro en 1931, mucho antes de que los científicos

descubrieran que es precisamente la capacidad de perdurar en el tiempo, más allá de los límites temporales convencionales, la característica fundamental de las moléculas que en el cerebro sostienen los recuerdos duraderos.

Es en las sinapsis, las bolsitas donde distintas neuronas se conectan, donde se almacenan las piezas básicas de un recuerdo, las distintas notas de la partitura de una sinfonía. Pero las moléculas que participan en la memoria se localizan por toda la neurona, también fuera de las sinapsis. Cuando una memoria se forma, una cascada de moléculas que se activan sucesivamente recorre el interior de las neuronas. Al principio, ciertas sinapsis se activan durante el aprendizaje, arrancando aquí una onda expansiva de actividades bioquímicas que se desplaza hasta el núcleo (otra bolsa dentro de la célula que contiene el material hereditario). Allí, ciertos genes se activan, dando paso a la fabricación de proteínas que volverán de nuevo a las sinapsis para modular su fuerza o incluso construir o destruir sinapsis enteras. Un verdadero “diálogo entre sinapsis y núcleo”. Así lo describió Eric Kandel (1929-), quien descubrió muchas de estas moléculas, al recibir el Premio Nobel en el año 2000 [7].

Las moléculas son la tinta de la memoria. Los recuerdos se escriben a fuerza de moléculas que envían, reciben o modulan la transmisión de mensajes entre neuronas. Algunas moléculas empaquetan los mensajes (el neurotransmisor), otras los reciben y convierten en impulsos nerviosos (los receptores), y otras afinan el instrumento para que se ajuste a una determinada tonalidad. Como en una cascada de piezas de dominó, sin embargo, la mayoría de esas moléculas, una vez hecha su actuación, dejan de ser necesarias para que el proceso continúe. ¿De qué forma entonces unas moléculas, que por otra parte duran horas o días, pueden almacenar recuerdos que duran toda la vida?

En 1984, Francis Crick (1916-2004), quien descubriera dos décadas antes la estructura del ADN, propuso una posibilidad para resolver el dilema. Crick pensó que alguna forma de actividad bioquímica que lograra permanecer tanto tiempo como los recuerdos podría ser la base de la persistencia de la memoria. Dos décadas más tarde, otros científicos descubrieron una molécula que no solo cumplía el requisito de perdurar, sino que además su función era esencialmente colocar receptores en la sinapsis y de ese modo potenciar las conexiones neuronales [8]. Todo lo que antes habían apuntado Cajal, Hebb y Crick cuadraba.

El nombre de esa molécula es PKMζ. No se trata de un extraño Pokémon griego, sino las siglas de una proteína llamada Proteína Kinasa M-Zeta. Una molécula que pertenece a un tipo de proteínas, las kinasas, de las que tenemos más de 500 en el cuerpo. PKMζ es especial entre todas ellas porque abunda en el cerebro y porque es capaz de perdurar durante años en las sinapsis que se activan al aprender.

Todas las kinasas son proteínas capaces de modificar a otras proteínas. Lo hacen añadiéndoles un pequeño trozo de molécula, como si añadieran la pieza que le falta a un puzle. Solo que en este caso, añaden justo la pieza donde está el botón de un interruptor, porque cuando una proteína es modificada, su comportamiento cambia de forma radical. En algunos casos, las proteínas se activan cuando son modificadas, como le sucede a los receptores de neurotransmisor. Una vez modificados por la acción de PKMζ, los receptores son entonces transportados al lugar donde pueden ejercer su función: al interior de las sinapsis. Y una vez allí en las sinapsis -esas bolsitas que actúan como centrales de telefonía enviando y recibiendo mensajes entre neuronas- los receptores son concretamente colocados en la membrana de la sinapsis. Justo el lugar donde pueden

detectar más neurotransmisor procedente de otras neuronas y potenciar la comunicación neuronal.

En otros casos, sin embargo, el significado del interruptor está invertido, de forma que la proteína es inactivada cuando es modificada por acción de PKMζ. Esto le sucede a otra proteína que actúa como una grúa quitando constantemente receptores de la membrana de la sinapsis. Estas grúas son unos verdaderos agentes del olvido que borran los trazos de tinta con los que se escribe la memoria en el cerebro. Pero al ser modificadas por PKMζ, estas grúas se bloquean, como si alguien colocara un obstáculo en las ruedas del sistema de poleas.

De esta forma, PKMζ contribuye a aumentar el número de receptores en las sinapsis de dos formas distintas: por un lado los añade y por otro lado evita que sean retirados. Por eso mismo, PKMζ es necesaria permanentemente para mantener la potenciación sináptica y que la memoria persista en el tiempo.

¿Y cómo logra PKMζ perdurar? Hay una tercera proteína que es su imagen antagónica. Es una proteína que se encarga de inhibir la producción de PKMζ. Y como en el caso anterior, esta proteína es inhibida cuando es modificada por PKMζ. Así que una vez presente, PKMζ elimina el freno a su propia fabricación. Así es como trasciende. Así es como puede preservar la tinta de las trazas de memoria.

81.4 No solo una molécula

Todo parecía cuadrar hasta que en 2013 dos laboratorios distintos realizaron una prueba de peso a la hipótesis de que PKMζ pudiera ser la molécula de la memoria [9]. Estos investigadores crearon ratones sin el gen de PKMζ, eliminando así la posibilidad de producir la proteína durante el aprendizaje. Uno esperaría que estos ratones fueran como la versión extrema de Henry Molaison, el paciente sin memoria, incapaces de formar cualquier recuerdo. Sin embargo, ninguno de los investigadores encontró consecuencias sobre la retención de memorias duraderas. Aquellos ratones tenían una memoria perfectamente normal.

Las hipótesis son barcos que navegan a través de duras tormentas. A veces se hunden como el Titanic, pero a veces aguantan el embiste furioso de gigantes olas e icebergs. En 2016, otro estudio descubrió una segunda molécula, muy parecida a PKMζ, que se produce cuando el gen de esta última es eliminado y que es capaz de sustituir su función. El barco aguantó esta vez, pero los científicos llegaron a una conclusión: PKMζ no es toda la tinta de la memoria. PKMζ no es el solista que protagoniza la sinfonía ni la única molécula que escribe las notas en el pentagrama de la memoria. PKMζ es parte de un gran sistema de moléculas capaces de retener los cambios en la comunicación neuronal que sostienen los recuerdos duraderos [10].

Hay al menos otras dos moléculas necesarias para formar memoria y que también logran mantener una actividad persistente más allá de lo convencional. Y con todo, ni siquiera estas cuatro moléculas actúan en solitario. Están todas ellas inmersas en una red de decenas de moléculas que rellenan las sinapsis como la tinta que escribe las notas de la partitura, grabando la sinfonía del lejano pero vivo recuerdo del primer beso. Un sistema de moléculas que, dentro de otro sistema de miles de neuronas esparcidas por el cerebro, comienzan a vibrar y embelesarse muy pronto tras el primer

contacto de labios. Labios que se funden, se acarician, que se atraen y empujan como las moléculas que en el cerebro interactúan y danzan. Un diálogo amoroso que inspira al hipocampo, un compositor escondido en un lugar recóndito del cerebro, a escribir la música de otro diálogo emocionante, vibrante, vital. Un diálogo hecho de diálogos de lenguas hechas de carne y neuronas hechas de moléculas hechas de átomos que recorren su interior como cascadas de dominó. Un diálogo de chispas que saltan entre centenares de neuronas que se afinan y sincronizan a un mismo tempo y tono, dentro de una orquesta que reproduce y que graba la obra musical que nunca podrán olvidar ninguno de los protagonistas.

81.5 Preguntas abiertas

Queda mucho por conocer sobre las bases biológicas de la memoria, así como de la reparación de la memoria en condiciones subóptimas.

¿De qué forma compone el hipocampo las memorias? ¿Es el único compositor en la orquesta del cerebro? Recientemente, tras la muerte de Henry Molaison, se han hecho estudios detallados de su cerebro, encontrando lesiones en otros lugares además del hipocampo [11]. En cualquier caso, sigue siendo un misterio cómo las memorias episódicas dejan de necesitar esta región cuando se hacen longevas.

También sabemos que hay distintos tipos de memorias. El propio Molaison tenía defectos en memorias episódicas, pero podía aprender tareas motoras. ¿Participan las mismas moléculas en memorias no declarativas?

A nivel molecular, el cuadro es todavía incompleto. ¿Cuántas moléculas hacen falta para construir un recuerdo? ¿Cuántas son suficientes? ¿Qué genes se expresan y qué proteínas deben fabricarse para inscribir una memoria a largo plazo?

Las moléculas persistentes, PKMζ y las demás, deben tener algún freno. De lo contrario las múltiples memorias que vamos adquiriendo con la vida acabarían saturando las sinapsis del cerebro. ¿Cómo se evita que esto ocurra?

¿Qué papel tiene la formación o la eliminación de sinapsis (cambios estructurales) respecto a la potenciación o debilitamiento de las sinapsis existentes (cambios funcionales)? ¿Es este equilibrio igual en todos los tipos de memoria y zonas del cerebro? ¿Qué papel exacto juega la potenciación respecto al debilitamiento de las conexiones sinápticas? ¿Y qué aspectos concretos de una memoria son almacenados por una sinapsis individual?

¿Cuántas redes neuronales codifican un recuerdo? ¿Hay redundancia, o cada red neuronal codifica un recuerdo? ¿Cuántos recuerdos puede una neurona almacenar? ¿Cuántas neuronas hacen falta para codificar un recuerdo?

Sabemos que en el cerebro hay muchos tipos de neuronas, que se distinguen, entre otras características, por el neurotransmisor o combinación de neurotransmisores que emplean (excitadoras, inhibidoras, etc.), por la longitud de sus proyecciones y por la zona del cerebro donde se encuentran. ¿Qué tipos de neuronas son necesarias para formar una memoria? ¿Cómo contribuye y qué almacena cada una de esas clases? ¿Qué

papel desempeñan las nuevas neuronas que continuamente se forman en el hipocampo? ¿Y qué funciones tienen otros tipos de células no neuronales, como la glía?

¿Cuánto dura en realidad una memoria? Algunos estudios recientes en ratón sugieren que algunos traumas aprendidos podrían transmitirse de una generación a otra. ¿Sucede esto de forma general, en otros tipos de memorias y en otras especies? Curiosamente, Richard Wolfgang Semon, el primero en hablar de la traza de memoria, fue un acérrimo defensor del Lamarckismo (la herencia de los caracteres adquiridos), e incluso pensaba que las memorias podían heredarse.

¿Cómo envejecen las memorias? ¿Permanecen intactas o van transformándose en el tiempo? La analogía del cerebro como un DVD que graba y reproduce los recuerdos no es del todo precisa: sabemos que la memoria tiene una escasa fidelidad. ¿Qué es entonces lo que persiste y cómo sucede ese proceso de transformación con el tiempo?

¿Cómo funciona el proceso de recordar? ¿Participan en él las mismas moléculas y mecanismos cerebrales que en la adquisición de memorias? Se sabe que durante la reactivación de los recuerdos tiene lugar un proceso llamado reconsolidación que permite modificar el contenido de los recuerdos [12], ¿cómo influye el proceso de recordar en la persistencia de la memoria? ¿Puede intervenirse en el momento de recordar una memoria para eliminar recuerdos traumáticos? ¿O para potenciar la persistencia de recuerdos que desaparecen, como ocurre en enfermedades neurodegenerativas?

Para que una memoria persista, tiene que resistir al olvido. ¿Cómo funciona este proceso? ¿Qué mecanismos cerebrales son responsables? ¿Puede frenarse el olvido para evitar la pérdida de memoria que ocurre con la edad, o acelerarse para borrar traumas? Hay niños y niñas con discapacidades intelectuales congénitas causadas por alteraciones en las moléculas que escriben en el cerebro la memoria a largo plazo. ¿Será posible algún día potenciar la memoria y el aprendizaje en un contexto educativo de una forma socialmente justa y que atienda a la vez a la diversidad de necesidades educativas?

¿Qué papel exacto tiene el sueño? Sabemos que el descanso es importante para formar memorias [13]. Algo que, años antes de que ningún científico lo describiera, el escritor Jorge Luis Borges (1899-1986) plasmó virtuosamente en el cuento "Funes el Memorioso" – una metáfora del insomnio que el autor padecía. El protagonista del cuento, Ireneo Funes, tenía una memoria portentosa. No era capaz de olvidar nada, todo lo recordaba. "Más recuerdos tengo yo solo que los que habrán tenido todos los hombres desde que el mundo es mundo", así describía el protagonista su capacidad. Pero al contrario de lo que uno pudiera sospechar, lejos de un don aquello se reveló como un auténtico calvario. No dormía bien y los recuerdos dolorosos lo martirizaban. "Mi memoria es como vaciadero de basuras", decía Funes. Todo detalle insignificante quedaba impregnado en su memoria. Incapaz de desechar detalles superfluos, Ireneo era incapaz de entender el mundo porque no podía generalizar. No podía, por ejemplo, saber si el mismo perro que había visto horas antes de perfil era el mismo perro que ahora veía de frente, porque ambas imágenes no coincidían exactamente. Quizá su calvario tuvo una única recompensa: la de guardar celosamente un recuerdo nítido, cristalino y puro, seguramente el más vívido jamás experimentado, de aquello con lo que todos soñamos de vez en cuando: el primer beso.

Bibliografía:
[1] E. D. Roberson and J. D. Sweatt, “A biochemical blueprint for long-term memory,” *Learn.Mem.*, vol. 6, no. 1072–0502, pp. 381–388, Jul. 1999.
[2] P. Pietikäinen, *Alchemists of human nature : psychological utopianism in Gross, Jung, Reich, and Fromm*. Pickering & Chatto, 2007.
[3] Rodrigo Quian Quiroga, *Borges y la memoria*. EDITORIAL SUDAMERICANA, 2012.
[4] R. G. Morris, “Elements of a neurobiological theory of hippocampal function: the role of synaptic plasticity, synaptic tagging and schemas,” *Eur.J.Neurosci.*, vol. 23, no. 0953–816X, pp. 2829–2846, Jun. 2006.
[5] B. Milner, L. R. Squire, and E. R. Kandel, “Cognitive neuroscience and the study of memory,” *Neuron*, vol. 20, no. 0896–6273, pp. 445–468, Mar. 1998.
[6] D. O. Hebb, *The Organization of Behavior*. New York: Wiley, 1949.
[7] E. R. Kandel, “The molecular biology of memory storage: a dialogue between genes and synapses,” *Science (80-.).*, vol. 294, no. 5544, pp. 1030–1038, Nov. 2001.
[8] T. C. Sacktor, “How does PKMζ maintain long-term memory?,” *Nat. Rev. Neurosci.*, vol. 12, no. 1, pp. 9–15, Jan. 2011.
[9] R. G. Morris, “Forget me not.,” *Elife*, vol. 5, 2016.
[10] P. W. Frankland and S. A. Josselyn, “Neuroscience: Memory and the single molecule.,” *Nature*, vol. 493, no. 7432, pp. 312–3, Jan. 2013.
[11] J. Annese, N. M. Schenker-Ahmed, H. Bartsch, P. Maechler, C. Sheh, N. Thomas, J. Kayano, A. Ghatan, N. Bresler, M. P. Frosch, R. Klaming, and S. Corkin, “Postmortem examination of patient H.M.’s brain based on histological sectioning and digital 3D reconstruction.,” *Nat. Commun.*, vol. 5, p. 3122, 2014.
[12] Y. Dudai, “The restless engram: consolidations never end.,” *Annu. Rev. Neurosci.*, vol. 35, pp. 227–47, 2012.
[13] R. Stickgold and M. P. Walker, “Sleep-dependent memory triage: evolving generalization through selective processing,” *Nat. Neurosci.*, vol. 16, no. 2, pp. 139–145, Jan. 2013.

José Viosca
Doctor en Neurociencias
Comunicador Científico / (**Twitter**)

Capítulo 82
¿Por qué se envejece? ¿Es el resultado de la aplicación de “la flecha del tiempo”?
(Por Alberto Quintero Gámez)

Antiguos filósofos afirmaron que la cesación de un fenómeno es producto de su surgimiento; todo lo que alguna vez nace esta condenado a permanecer y morir. Dado que el conocimiento en el ser humano es adquirido de una forma dualista (porque es como se le presenta el universo), solo a través del surgimiento se puede definir la cesación y solo a través de la cesación se puede definir el surgimiento. Por lo tanto la degeneración que presenciamos en nuestra propia vida al pasar los años no solo nos pertenece a los seres vivos, en realidad aplica de forma natural a todos los fenómenos que nos rodean, los edificios cambian y desaparecen, al igual que las ciudades, montañas, océanos, planetas, galaxias y todo el universo en un constante movimiento sin descanso.

Al margen del profundo carácter filosófico que puede tener el plantearse preguntas sobre el declive inherente en todos los fenómenos que nos rodean, partiendo de esa intriga generada por los filósofos como sus predecesores, la ciencia en la actualidad dominada por una corriente de pensamiento más materialista, ha intentado estudiar que ocurre físicamente en los organismos que consideramos seres vivos que produce su desaparición. Los organismos vivos son maquinarias muy complejas al compararlas con las desarrolladas por los seres humanos, tienen componentes muy variados y funciones muy complejas que dependen de una gran cantidad de interacciones.

Nuestros coches son maquinas bastante simples en comparación con nuestros organismos, al fallar alguna de sus partes o tener un daño grave en su estructura el coche dejara de funcionar y se podrá desmontar para usar de nuevo los materiales que le conformaron, de forma similar nuestro organismo cuando se encuentra con el daño de algún componente o sufre daños generalizados se detiene y el material que lo conforma es transformado con el paso del tiempo por el entorno. Pero a diferencia del coche que se detiene cuando se presenta un fallo, los seres vivos parecen tener un límite a su propia vida aunque ningún problema se presentase, las células parecen estar programadas para transcurrido un tiempo especifico detener su funcionamiento. ¿Podría esto tener algún sentido cuando el deseo de todas las especies es seguir existiendo? ¿Qué propósito puede tener el envejecimiento? ¿Puede ser beneficioso?

Es evidente que una de las características de los seres vivos es su diversidad, existe una variedad enorme de pequeños organismos e incluso en la actualidad hay siete mil millones de humanos todos con características muy diferentes los unos de los otros. Esta diversidad se genera debido a que en el proceso de división celular ocurre de forma constante. Pequeños cambios en el código que define lo que será ese organismo, son cambios aleatorios que en ocasiones pueden ser útiles para la existencia del organismo y

en otras ocasiones no tanto, generando de esta forma variedad. El intercambio de información genética entre organismos diferentes puede acelerar el ritmo con el que se produce esta diversidad.

Los nuevos organismos con ligeras diferencias respecto a sus progenitores podrían tener mejores características para relacionarse con su entorno superando obstáculos mientras que otros no tanto, y es evidente que sería muy útil para el progreso de todo el conjunto que los que no tuviesen tanto éxito pudiesen retirarse de la escena de forma de hacer más espacio para los nuevos organismos mejorados, obviamente los menos aptos desaparecerían al igual que el coche por algún desperfecto de una de sus partes o un daño generalizado pero sería mucho más eficiente acelerar el proceso.

Por lo tanto aquellas especies con periodos de vida más cortos tendrían más diversidad y por lo tanto mayor oportunidad de conseguir cambios que les fuesen útiles. Una evidencia clara de lo enunciado anteriormente son las moscas que tienen periodos de vida tan cortos como una semana y existen cerca de 150.000 especies, mientras que en las hormigas que viven entre 1 mes y 3 años existen cerca de 15.000 especies, finalmente dentro del grupo de animales más longevos como la medusa "Turritopsis nutricula" que es famosa por revertir el proceso de envejecimiento una vez alcanzada la madurez sexual y rejuvenecer de forma que vive prácticamente hasta ser devorada por un depredador, pertenece a la clase Hydrozoa dentro de la cual hay unas 3.000 especies.

Investigaciones actuales han comenzado a entender como este complejo proceso de envejecimiento inducido en los organismos es posible. Se han identificado secciones en los extremos de los cromosomas denominadas telómeros que tienen la función de preservar la estabilidad estructural de los cromosomas, a partir de este conocimiento se han generado teorías sobre el envejecimiento adjudicando a estos telómeros el papel de reloj celular al definir el número de divisiones celulares posibles hasta que la célula muere. En el proceso de división celular estos telómeros se van haciendo cada vez más cortos, hasta alcanzar un límite en el cual no pueden desempeñar su papel de protector de la estructura del cromosoma y se interrumpe la división celular, las células no solo son incapaces de reproducirse sino que activan el proceso de apoptosis o muerte celular programada. Esta apoptosis tiene una función esencial en el desarrollo de la vida, evitando problemas como el cáncer que es la reproducción de forma indiscriminada de una célula ya dañada por medio de la producción de telomerasa una enzima que evita el acortamiento de los telómeros, otro ejemplo es la diferenciación de los dedos humanos durante el desarrollo embrionario por la apoptosis de las células de las membranas intermedias. Por lo tanto aunque resulte un poco irónico la muerte celular programada es causa de vida.

Finalmente aunque en nuestras pequeñas mentes entendamos el universo como un conjunto de fenómenos discretos la realidad es un continuo sin fronteras en constante movimiento, investigadores como Alex Wissner Gross proponen que la capacidad de un ente para desarrollar inteligencia y ser consciente de su entorno es producto de una maximización de la entropía o grado de desorden del mismo, solo los sistemas altamente entrópicos pueden producir seres que son capaces de ser conscientes del entorno. Este estado de máxima entropía estará favorecido por el hecho de que hayan muchos surgimientos y cesaciones en periodos más cortos de tiempo, por lo que el envejecimiento acelerado traerá un beneficio evolutivo dando la posibilidad de generar seres tan maravillosamente complejos como lo somos los seres humanos, siendo

capaces de experimentar la vejez, poder preguntarse porque esto ocurre y no detenerse hasta encontrar una respuesta.

Si el tiempo es solo una abstracción para expresar lo que observamos o es una variable con carácter físico que efectivamente puede ser modificada de forma de viajar en ambas direcciones, no solo en un línea recta apuntando al futuro, es un tema de la física más que de la biología, de poder mover el tiempo en dirección inversa los fenómenos ocurrirán en dirección inversa, todos los hechos ocurrirían también hacía atrás, la evolución de los organismos incluida dentro de ellos. Es mejor que haya avance aunque eso implique envejecer.

Alberto Quintero Gámez
Magister Matemática Industrial
Investigador en Universidad Carlos III de Madrid

Apéndice

Incluimos este apéndice como complemento a la explicación del ¿por qué? de este libro. Una gran mayoría, sino todos, de los autores participantes, en algún momento, han disfrutado con la lectura de algún libro de Isaac Asimov. Leyendo la introducción de *Cien preguntas básicas sobre la Ciencia*, creímos que era necesario, en estos momentos de tanta especialización, hacer una especie de "revisión / actualización / homenaje" de este clásico de la divulgación científica.

Adjuntamos la introducción e índice del original. Tras 50 años de ciencia los temas candentes e interesantes han variado, hemos intentado que este libro sea un buen complemento a la lectura del "cincuentón". Creemos que la lectura de los libros de divulgación, incluso los de ciencia ficción también, de Isaac Asimov es una muy buena forma de mantener viva la "curiosidad del niño pequeño" que todos llevamos dentro.

“””Introducción – Cien preguntas básicas sobre la Ciencia

Por el año 1965, esa respetable revista que es Science Digest inició una nueva sección titulada "Please Explain" ("Por favor, explique"). El propósito de esta sección era seleccionar algunas de las preguntas formuladas por los lectores y contestarlas en unas 500 palabras. La revista me preguntó si estaba dispuesto a abordar de vez en cuando una pregunta a cambio de una cantidad razonable de dinero. "Bien", contesté con ciertas reservas, "pero siempre que sea de vez en cuando". Tuve que suponerlo. Mi colaboración, que en principio iba a ser esporádica, adquirió carácter mensual, y la sección "Please Explain" pasó a ser "Isaac Asimov Explains" ("I. A. explica"). (Para evitar la posible trampa de mi archiconocida modestia diré que el cambio se hizo sin consultarme.) Cuando quise darme cuenta, llevaba ya colaborando más de ocho años y había acumulado un centenar de preguntas y respuestas. ¿Y quién podría resistir la tentación de reunir esos ensayos y hacer con ellos un libro? ¡Yo, desde luego, no! ¡Ni tampoco Houghton Mifflin! Como las respuestas que tengo que dar dependen de las preguntas que formulan los lectores, los ensayos no se reparten uniformemente por todo el campo de la ciencia. Por alguna razón, los lectores se hallan profundamente interesados por la física teórica, siendo especialmente numerosas las preguntas acerca de la velocidad de la luz y de las partículas subatómicas. De ahí que exista cierta duplicidad entre las respuestas, así como algunas omisiones flagrantes. Ambas cosas tienen sus ventajas. Las duplicidades provienen, en parte, que he intentado hacer las respuestas lo más completas posible. El lector puede consultar por tanto cualquier cuestión que se le venga a la imaginación y leer el libro en el orden que le plazca. Y en cuanto a las omisiones... ¿qué hay de malo en dejar que susciten una sana curiosidad? Si es suficientemente sana, envíe su propia pregunta a Science Digest. Si tengo ocasión (y sé lo suficiente) la contestaré, y de aquí a ocho años podría haber material bastante para publicar un libro titulado "Isaac Asimov sigue explicando".

Indice – Cien preguntas básicas sobre la Ciencia

44 *¿Qué es la antigravedad? ¿Cómo puede estudiarse?*
45 *¿Cuál es la velocidad de la gravitación?*
46 *¿Qué es la teoría del campo unificado?*
47 *¿Qué es, en pocas palabras, la teoría de la relatividad de Einstein?*
48 *(Parte 1) ¿Por qué la materia no puede moverse más deprisa que la velocidad de la luz?*
49 *(Parte 2) ¿Por qué la materia no puede moverse más deprisa que la velocidad de la luz?*
50 *Las partículas que se mueven más deprisa que la luz emiten radiación luminosa. ¿Cómo es posible, si no hay nada que se propague más deprisa que la luz?*
51 *Si no hay nada más rápido que la luz, ¿qué son los taquiones, que al parecer se mueven más deprisa que ella?*
52 *Los taquiones de energía cero se mueven con velocidad infinita. ¿Es de verdad posible una velocidad infinita?*
53 *¿Qué es el principio de incertidumbre de Heisenberg?*
54 *¿Qué es la paridad?*
55 *¿Por qué se habla de la vida media de un isótopo y no de su vida entera?*
56 *¿Por qué están encontrando los científicos tantas partículas subatómicas nuevas y cuál es su significado?*
57 *¿Qué es un quark?*
58 *Se ha dicho que los protones están constituidos por combinaciones de tres quarks p y también que un quark es treinta veces más pesado que un protón. ¿Cómo pueden ser ciertas ambas cosas a la vez?*
59 *En la bomba atómica se convierte materia en energía. ¿Es posible hacer lo contrario y convertir energía en materia?*
60 *Las antipartículas ¿producen antienergía?*
61 *¿En qué difieren las propiedades de los rayos cósmicos de las de los neutrinos?*
62 *¿Qué peligro encierran los rayos cósmicos para los hombres en el espacio?*
63 *Los neutrinos ¿son materia o energía?*
64 *¿Cómo funciona una cámara de burbujas?*
65 *¿Qué es un reactor generador?*
66 *¿Cuánto y durante cuánto tiempo hay que calentar el hidrógeno para mantener una reacción de fusión?*
67 *¿Cómo funciona un microscopio electrónico?*
68 *¿Qué es la entropía?*
69 *¿Está degradándose el universo?*
70 *¿Qué relación hay entre entropía y orden?*
71 *¿Qué relación hay entre la entropía y el tiempo?*
72 *Si el universo está constantemente degradándose, ¿cómo fue al principio?*
73 *Las ondas de radio y las ondas luminosas se utilizan para "ver" cosas en el espacio. ¿Hay otras clases de ondas con las que podamos "ver"?*
74 *Al calentar una sustancia se pone primero roja, luego naranja, después amarilla, pero a continuación blanca. ¿Por qué no sigue el espectro y se pone verde?*
75 *¿Qué es la luz polarizada?*
76 *La luz ¿puede ejercer fuerza sobre la materia?*
77 *La luz roja es la menos desviada al pasar por un prisma, pero la que más se desvía al pasar por una red de difracción. ¿Por qué esa diferencia?*
78 *¿Qué ocurre con la energía cuando dos haces de luz interfieren y producen oscuridad?*
79 *¿Qué es el efecto Coriolis?*
80 *El sonido se mueve más deprisa en sustancias densas como el agua o el acero que en el aire; sin embargo se mueve más deprisa en el aire caliente que en el frío, cuando el aire caliente es menos denso que el frío. ¿Es una paradoja?*
81 *¿Se hunden los barcos hasta el fondo del mar o llega un momento en que la presión les impide seguir bajando?*
82 *¿Cuáles son los elementos químicos más activos y por qué?*

83 ¿Qué tienen de noble los gases nobles?
84 ¿Por qué se forman los cristales y por qué lo hacen siempre en ciertas formas?
85 ¿Se puede comprimir el agua?
86 ¿Qué es el hidrógeno metálico? ¿Cómo puede ser el hidrógeno un metal?
87 ¿Qué es la "poliagua"? Si sigue siendo H2O, ¿cuál es la diferencia?
88 ¿Por qué se dilata el agua al congelarse?
89 ¿Qué son las pilas de combustible? ¿Qué ventajas presentan en la generación de electricidad?
90 ¿Qué son las vitaminas y por qué las necesitamos?
91 ¿Cómo empezó la vida?
92 ¿Es posible una vida de silicio?
93 ¿Por qué se extinguieron los dinosaurios?
94 ¿Cuál es la diferencia entre un cerebro y un computador? ¿Pueden pensar los computadores?
95 ¿Cuál es la velocidad del pensamiento?
96 ¿Qué son los "relojes biológicos" y cómo funcionan?
97 ¿Cuál es la diferencia entre bacterias, microbios, gérmenes y virus?
98 ¿Cómo se descubrieron los virus?
99 ¿Por qué las células de la sangre se reponen cada pocos meses, mientras que la mayoría de las células del cerebro duran toda la vida?
100 ¿Qué fin tiene el envejecer?

Cien preguntas básicas sobre la Ciencia – Isaac Asimov *“””*

Ruegos y preguntas

Tras la lectura de este libro, ya sea al completo o algún capítulo individualmente, me imagino, o eso espero, que tendréis multitud de preguntas sobre los temas tratados.

Diréis: ¿y ahora a qué viene esto?

En el momento de escribir estas líneas acabo de hacer una primera lectura del capítulo 51, ¿Qué son los neutrinos? Y tengo en la cabeza, cual estribillo de una canción pegadiza, la duda o sencillamente la pregunta de: si los neutrinos prácticamente no interaccionan, ¿por qué los procedentes del Big Bang (con lo que tuvo que ser aquello) son menos energéticos que otros?

Casi todas las preguntas, o dudas, se contestaran con la lectura más a fondo de la bibliografía recomendada por cada uno de los autores.

Pero si creéis que tenéis la pregunta crucial, esa que os obligaría a levantar la mano y preguntar al final de una clase o al final de una conferencia, no dudéis en enviármela:

quintingarrido@gmail.com

Intentaré, que en un tiempo más o menos breve, poder enviaros una respuesta.

Tal vez, y solo tal vez, con esa colección de preguntas y respuestas se podría plantear el proyecto de llevar a cabo "otra idea feliz" y agruparlas dando lugar a una especie de 2ª parte de este libro que estáis a punto de llegar a su final.

Esperando que la lectura de este libro os haya sido un agradable placer y sintáis que "estáis a hombros de gigantes".

Un abrazo.
Quintín

Quintín Garrido Garrido

Madrid – 1969

Con buenas aptitudes para la Física y las Matemáticas, este contable desde comienzos de la década de los 90, cursó sus estudios en el I.E.S. Cervantes de Madrid e inició los de Ciencias Físicas en la Universidad Autónoma de Madrid, no terminándolos.
Tirador con arco, montañero, navegante de barcos a vela (de radio control) y sobre todo lector de divulgación científica en general y de Física en particular.

Creador de dos blogs con tintes científicos:

http://momentosestelaresdelaciencia.blogspot.com.es/
Biografías de científicos en castellano. Desde Enero de 2013.

Y

http://divulgacioncientificadecientificos.blogspot.com.es/
Fundamentalmente recomendaciones de libros de divulgación, hechas por científicos, en castellano e inglés. Desde Noviembre de 2014.

En este libro consigue aglutinar a muchos de los científicos que le han recomendado libros en su blog y los pone a "divulgar" sobre cuestiones básicas de la ciencia.



Colaboran en esta edición:

- Alberto Aparici
- Alberto Casas
- Alberto González Fairén
- Alberto Quintero Gámez
- Ana Inés Gómez de Castro
- Angel M. Uranga Urteaga
- Angel S. Sanz
- Antonio Guerrero Ortega
- Bartolo Luque
- Belén Valenzuela Requena
- Bernardo Herradón
- Carlos M. Pina
- César González Arranz
- Christine Allen
- David L. Van Vranken
- Elena Denia
- Emilio García Ladona
- Enrique Macías Virgós
- Enrique Zuazua
- Felipe Zapata Ruiz
- Francisco J. Terán
- Gerardo Ortiz
- Ignacio Durán Escribano
- Ignacio Negueruela Díez
- Isabel Caballero de Frutos
- Joaquim Ballabrera Poy
- Jorge Pla-García
- José A. Caballero
- José Edelstein
- José Luis Fernández Barbón
- José Luis Rubio
- José Manuel Llorens Montolio
- José Miguel García Martín
- José Oñorbe Bernis
- José Ramón Martínez Saavedra
- José Viosca
- Juan Antonio Aguilar Sánchez
- Juan García-Bellido Capdevila
- Julio Gutiérrez Muñoz
- Lance Fortnow
- Laura Bayés
- Luis Felipe Rodríguez Jorge
- Luis Julián Goicoechea Santamaria
- Luis Velilla Prieto
- María A. H. Vozmediano
- Mario Sánchez Sanz
- Marta Macho Stadler
- Mayte Conejero Muriel
- Miguel Peiró
- Miguel Zumalacárregui Pérez
- Miquel Àngel Cuevas-Diarte
- Olga Caballero Calero
- Pablo López Tarifa
- Pablo Marcos Arenal
- Paola Marziani
- Pedro Alegría Ezquerra
- Pedro A. Pereda
- Raghuveer Parthasarathy
- Réne Duffard
- Ricardo David Fernández Cruz
- Ricardo Dorda Laforet
- Teresa Calvet
- Vanessa Arredondo
- Víctor Martín Lozano
- Xavier Hernández Döring
- Yaiza M. Castillo de la Peña
- Yves Huttel

